May 2016

UNDERSTANDING SOFT MOLECULAR INTERFACES USING MULTISCALE MOLECULAR DYNAMICS

Xiaoyi Wang
Syracuse University

Abstract:

Molecular dynamics simulations have proven to be a breakthrough approach in providing molecular structure and function details of biological macromolecules. The fact that molecular simulations can provide details of individual particle motions as a function of time, opens infinite possibilities of answering specific questions about a model system with atomistic and molecular precision, often more easily than experiments on the actual system. Instead of relying solely on experiments, molecular dynamics systems are based on empirical force fields that mimic molecular interactions over time. There has been increasing interest in utilizing the multiscale molecular dynamics (MD) approach to simulate the properties of biological macromolecules for meaningful applications in areas of nanotoxicity, nanotherapy, and nanomedicine.

Recent advances in nanomedicine have led to the great development of several drug-delivery platforms for targeted delivery. Polymeric drug delivery systems are designed and applied to ameliorate undesirable properties of the drug agents such as hydrophobicity, poor targeting ability, and broad scale toxicity. However, rational structural design of polymeric nanocarrier is only based on theoretical investigation, that restraint by absence of real molecular-level details. So, MD approach is used to provide those desirable details which not feasible in experiments. In our work, models of three generations of anticancer nanocarriers polymer were developed and characterized using multiscale MD simulations. The results were compared to *in vitro* experimental results. Observations were analyzed and found to be in good agreement with experiments results on size and morphology changes. Details shown by our systems helped us discover the reasons for the different nanocarriers' performances. Our results show the *in silico* methods that can be used to contribute to drug nanocarrier design optimization work.

Claudins, are critical components in building tight junctions (TJs) which could form paracellular channels or barriers for physiological functions. Recently, 27 types of claudins have been classified based on different functions and characteristics, and they are becoming potential points in developing drug delivery systems and pathological studies. However, the architecture of claudin strand in TJs is still unclear thus impeding further study and applications. So, we use MD simulation approach to replicate self-assembly process and try to find out the potential construction models. In our work, claudin-1, -2, -5, and -15 monomer models were built by homology modeling and validated. Self-assembly processes of non-mutated and mutated claudins were conducted to reproduce the construction of claudin macromolecular strands. Four classified dimer models predicted by existing research were found to be reproduced in our molecular dynamics systems, and their number distributions were calculated. Our results first time showed the potential models for TJ architectures that can be a guidance for understanding TJs.

UNDERSTANDING SOFT MOLECULAR INTERFACES USING MULTISCALE MOLECULAR DYNAMICS

By

Xiaoyi Wang

B.S. Shanghai Ocean University, 2013

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

in Biomedical Engineering

in the graduate School of Syracuse University

May 2016

Acknowledgement

First, I would like to thank my advisor, Dr. Shikha Nangia, who gave me the golden opportunity to join the lab and participate in several projects. With her constant support and guidance, I have learned so much fantastic high-end techniques on molecular dynamics which I haven't touched before.

Second, I would like to thank all my lab mates, Wenjuan, Jerome, Huilin, Nan, for their help in both research and life. I sincerely appreciate their support throughout my time in Syracuse. Our friendship will last forever.

Last but not least, I would like to thank my family. Thank all of them for supporting me on both emotion and finance. Without their support, I will never have such a valuable experience in such a great university.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Objective

Molecular dynamics (MD) simulations have emerged as a powerful approach to study physical motion of atoms and molecules as a function of time[1]. This approach can provide dynamical information about the system in atomic level detail, which is often not feasible via experiments[2]. Due to the inherent complexity of the biological system, computational methods have become increasingly popular to study biochemical systems including biocompatible polymers, drug-delivery nanocarriers[3], membrane proteins[4], enzymes, nucleic acids[5] among others. In recent times larger and more complex chemical systems are being investigated using MD due to availability of affordable and powerful computers.

In this work, MD simulations have been employed to study two unique chemical and biological systems. In the first system, optimization of anticancer drug-delivery nanocarriers is investigated for Food and Drug administration (FDA) approved drug Doxorubicin[6]. In the second system, self-assembly of Claudin which is a kind of membrane proteins responsible for the tight junctions between neighboring cells through the body[7], is investigated to understand the molecular architecture of the ubiquitous tight junctions.

It is well-understood that efficacy of a biocompatible polymeric nanocarrier relies on both the drug payload and the polymeric scaffold encapsulating the drug[8]. Different types of building blocks interact with drug molecules differently resulting highly variable physiochemical properties such as size distribution, shape, targeting specificity, and stability[9]. Small differences in molecular structure of building block can potentially have significant effect on drug-encapsulation and

micellization[10]. Therefore, examining the molecular interactions between drugs and polymeric scaffold is critical for nanocarrier design. In first chapter, self-assembly processes of three generations of linear-dendrimer copolymers with FDA approved drug doxorubicin are presented. Morphologies of nanocarriers formed in MD simulations were compared to the experimental data and the results are presented.

In the second chapter, we examine the *cis*-interactions of claudin proteins in lipid membranes. In general, claudin family of proteins form tight junctions complexes in the space lining two neighboring epithelial or endothelial cells to regulate intercellular permeability in underlying tissue[7], however the molecular origin of the architecture was not revealed clearly[11]. With better understanding of the relationship between structure and function of tight junctions, we will likely be more closer to find cures for diseases caused by the presence or absence of tight junctions, such as hepatitis C virus[12], Alzheimer's[13] and Parkinson's[14]. In order to understand the molecular architecture of the tight junctions, we adopted a systematic step-wise approach. To start monomeric structure of claudin, protein was built by homology modeling methods[15] and validated using a variety of computational tools[16]. In the second step, the claudin monomers were embedded in lipid membrane and optimized to obtain equilibrated structures. In the following step, cis-assembly of claudin proteins over tens of microseconds was performed that led of oligomeric strand formation. Close analysis of the dimeric interface in the strands that consistently led to four unique dimeric interfaces. All four of these dimers have been reported in the literature. The direct agreement between experimental and simulation results validates our computational approach and is promising for the research currently underway in our group on this topic.

1.2 Molecular Dynamics Simulations

Molecular dynamics simulations have been used in numerous studies to understand the physical basis of the structure and function in biological systems since the late 1970s[1-2]. To complement the experiments, atomistic models[4] are created to mimic the real system that are then propagated in time to reveal molecular-level details of the system. The movement atoms and molecules are governed by the Newton's equation, and the interactions between bonded and non-bonded atoms provide by the underlying force field. Several force fields such as AMBER[17], GLYCAM[18], CHARMM[19], and GROMOS[19b] have been developed to provide experimentally comparable structural properties in atomistic detail. In general, force field is defined as the sum of bonded and non-bonded interactions. Bonded interactions include bond-stretching, angle-bending, and dihedral terms, and non-bonded van der Waals interaction via the Lennard-Jones potential and electrostatic interactions using the Coulomb's law[20]. The reference value in bonded interaction determine the ideal structural relation among the atoms within one molecules and control the energetic terms. The non-bonded interaction considers the atoms overlap and dispersion and it is responsible for calculating the atomic repulsion and attraction among the different molecules. Based on force field and Newton's equation, energy and potential was considered and positions and velocities would be calculated in each time step, and then atomistic models have ability to interact and move within a period of time. As all positions are calculated with in timescale, motion trajectories of atoms and molecules could be shown.[2]

1.3 Coarse-grained model

MD applications often face problems about time-scale and length-scale disparity between computational and experimental methods[21]. Larger time-scale and length-scale computational works could provide more

means to represent nature and would be great help to experimentation. As longer and more complicated system simulations are required, atomistic models become a somehow too luxury approach for computational works[22]. Since, all atom-atom interaction are being calculated during the period of time, the simulation times are largely limited by the volume and complexity of system. And more calculations within the system leads to burden on computational resources. Practically, atomistic models are usually limited to nanosecond scale simulation times and around 10 nanometers system sizes which cannot meet the demands in current scientific study.

In order to better bridge computational to experimental methods, an alternative approach to overcome limitations of atomistic simulations is to employ coarse-grained (CG) models[23]. CG models are derived from atomistic model where on an average of four heavy atoms are assigned to one bead, resulting in decrease the number of particles in the system. The reduced number of particles along with larger integration step size in CG simulations leads to increase in improved efficiency of CG simulation by three or more orders of magnitude. Although the CG model decreases the resolution of atom-atom interactions, thermodynamic and structural properties of molecules could be reproduced and compromised by force field in CG model. One of popular CG force field that is frequently used is the Martini force field. In this model, each CG model bead is assigned in four types of interaction of sites, polar (P), nonpolar (N), apolar (C), and charged (Q), and each type has four to five subtypes to include subtle differences in chemical properties such as hydrogen-bonding abilities and polarity[23b, 24].

With the development of parameterization, various CG model have been developed to represent similar geometries and properties as atomistic model to help build more realistic and complex systems, such as solvent, ions, phospholipids, ring particles, and antifreeze particles[23b]. Even though CG models remove some chemical and spatial resolution and apply modified kinetics to

these beads, it could provide same overall movements and interactions of molecules as atomistic models for longer time. Introduction of more mature CG models benefits current biological mechanisms study a lot. CG models have been widely used in lots of areas in last decades, like biomolecules simulation including lipid, enzyme, protein and etc. In summary, utilization of CG model is beneficial for exploring more complex system in a larger timescale.

1.4 Application of MD

MD simulation approach has been used to both nanocarriers' design and membrane protein areas as same as the works in this thesis, and it has been proven to be capable to help experimentation. There are several examples listed as following.

Molecular and atomistic level details in designing nanocarrier are hard to be tracked because of the complexity and sophistication in experimentation. And MD approach could help reveal more unfeasible details in experimentation that could help designing the anticancer nanocarriers. A fair number of molecular dynamics cases have been done previously. Sharon M. Loverde et al, used rational coarse grain models to study PEG–PCL copolymer performance on anticancer drug Taxol[25]. By adjusting ratio of PEG and PCL in computational ways, relationships among micelle morphology, hydrophilicity, diffusion, transfer energy and performance were built properly which the process is hard and resources-consuming in experimentation. Another example in MD application in drug nanocarrier is, Mengyao Zheng et al studied siRNA delivery efficiency of PEI which is the most widely used polyction for nucleic acid delivery and its triblock copolymer. In that work, atomistic details of self-assembly of polyplexes with siRNA versus pDNA was shown by molecular dynamics ways which is a blind spot in experimentation. By using synergic method with experimentation, they claimed a relationship between nitrogen to phosphate ration, toxicity and transfection efficiency[26]. Both these two examples show applications of MD simulations could

help nanocarrier design work in a meticulous way that make copolymer to be more easily modified and evaluated.

MD simulations are also used to synergizes experimentation works explore the role of specific amino acid in protein folding process or biological functions. Most of experimentations use mutagens approach that compare the functions, morphologies, and functions between non-mutated and mutated system[27]. However, complexity in experimentation lead to loss of atomic and molecular resolution. MD simulations could detect detailed macromolecular constructions, track the atomistic interaction and calculate energetics in the biological system that might help reveal biological mechanisms, which these works could not be achieved by experimentation work. Therefore, molecular dynamics computational approaches becomes increasingly important in biophysical and biological mechanism study. In claudin area, there has been some molecular dynamics studies in last 3 years. Bhavaniprasad Vipperlaa et al used molecular dynamics approach to study claudin-1 point mutations effect in HCV infection[28]. By comparing several mutant models with native models, they claimed the single mutants would have unnoticeable change while double mutants models would affect the protein structures which the observation matched experimental works. In the same year, they use the same approaches to clarify Lys65 in claudin-4 as the potential amino acid point to claudin-4-ephrin type-A receptor 2 binding[29].

As these examples shown above, molecular dynamics computational approaches could be used to study both nanocarrier and protein folding/assembly areas. And this thesis will use molecular dynamics computational approaches to help understanding self-assembly processes of polymeric nanocarriers and tight junction proteins.

2 Designing anticancer drug delivery nanocarriers

2.1 Introduction

Cancer has been the world's most fatal problem for long periods. Over 10 million cases are diagnosed and millions of patients die every year, and this number is still ascending increasingly[30]. However, until now, although no perfect solution was invented to solve this problem, better understanding of cancer mechanism, higher medical technology and mature treatment skills were performed to solve that problem[31]. In recent years, chemotherapy treatment is one of the most used approaches to therapy cancer due to its widespread treatment range, and various types of natural or synthesized drug were found to have great efforts for anticancer treatments[31d, 32]. Despite the great effort to kill cancer cells, they have defaults that obstruct the application such as hydrophobicity that cannot be easily delivered in body environment, poor target ability that would be wasted during being transported to other area, the toxicity that will cause damage to other good organisms[33]. For example, Doxorubicin (DOX) was a kind of the main stream FDA-approved anticancer drug on the market to be used frequently for leukaemia and lymphoma cancer treatment[34]. DOX has been testified to have excellent performance on anticancer treatment, but its poor cancer-targeting ability and poor intratumoral leads to higher dose usage and cause life-threatening cardiotoxicity[35]. The adverse effect largely limits the dosage to be used thus the applications were restricted. Another widely used FDA-approved anticancer drug is Paclitaxel. It has been used in anticancer chemotherapy for ovarian, breast and lung, bladder and etc for more than ten years. However, hydrophobicity, low bioavailability and rapid drug resistance effect[36] obstruct the usage of Paclitaxel[37]. So, lots of effort has been devoted to develop nanotechnology to help improve efficiency and solve the defaults of those drugs.

Nanocarriers are nano-scale size material used for transport substances to the target places[8b]. It is commonly used in medicine area for cancer treatment. Because of it is biocompatible, small sized and designable, nanocarrier has become a potential method of solving the problems mentioned above. Several aspects are critical and challenging in determining nanocarriers' performances such as stability, drug-loading capacity, size distribution, and target specificity[38]. In drug administration, drugs are usually be delivered by oral, injection, ophthalmic, infusion, or intramuscular[39]. So, nanocarriers are required to have ability to protect the effective compounds in both chemical and physical area such as acid environment, blood stream shear strength, and high pressure. Also, nanocarriers need to have a good drug-loading capacity that they can carry more active compounds within the same dose and have a better effect. Size distribution is another important part related to biodistribution properties. Generally smaller size nanocarriers could more easily penetrate and have a higher cell uptake rate. And narrow size distribution contributes to stable drug release ratio and design improvement work. Active drug compounds usually have poor target ability, and they will infuse in body environment without specificity which will result in great waste, resistance and toxicity. Well-designed nanocarrier could improve the drug in chemical and physical stability in body environment, effectiveness, efficiency and biocompatibility[40].

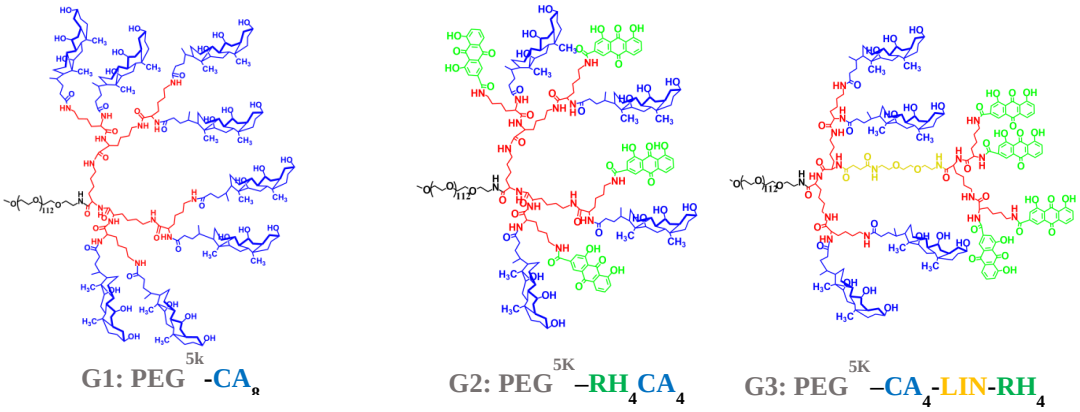


Figure 1 Structures of three genderations of Telodendrimer. Black parts: Polyethylene glycol (PEG); Red parts: Lysine; Blue parts: Cholic acid (CA); Green parts: Rhein (RH); Gold parts: link (LIN)

Block copolymer is a promising way to be used as nanocarriers because its designability and diversity[8a]. Several functional building blocks are copolymerized together to solve the challenges discussed above. With the amphiphilic character of block copolymer, hydrophobic active compounds will be automatically self-assembled and encapsulated within the copolymers, which is also easy to produce. Our work is based on a novel poly (ethylene glycol) (PEG) linear-dendrimeric amphiphilic block copolymer, $PEG^{5k}CA_8$, so-called telodendrimer(Telo) which is developed by Dr. Luo's research group[41]. $PEG^{5k}CA_8$ telodendrimer is a linear copolymer built by 112 polyethylene glycol (PEG), 7 lysine (LYS), and 8 cholic acid (CHOA) as shown in Fig1 and it has been claimed to have great performance as a nanocarrier for hydrophobic drug paclitaxel. It can enhance blood compatibility and accumulation ability at tumor based by its slightly negative charge micelle surface. Recently, in order to extend $PEG^{5k}CA_8$ Telo to be used on new drug doxorubicin, two new generations of telodedrimer were invented by replacing building blocks on the structural platform. Four CHOA were replaced by rhein (RHN), a short amphiphilic link (LIN) was added and building blocks were reorganized to second generation Telo $PEG^{5K}Rh_4CA_4$ and third generation Telo $PEG^{5k}CA_4$–LIN-Rh_4[42]. Luo's group proposed introduction of rhein group will help self-assemble process with doxorubicin by pi-pi stacking and high binding energy. These two generations telodendrimer were validated to be able to self-assemble with doxorubicin and form micelles, the performances between these two generations telodendrimer were compared and discussed. The results showed that second generation telodendrimer has lower drug release ratio than that of third generation which is better in clinical use. Also, the shape of third generation telodendrimer micelle is predominately longer worm like shape that make size distribution wider while second generation telodendrimer micelle forms shorter worm like shape micelle. It was somehow out of their expectation that third generation Telo was expected to have more stable

interaction with doxorubicin because all binding rhein group in one side might provide more stable interactions, while in the second generation RHN building blocks are paired with CHOA that might have lower stable interaction.

In order to further study drug delivery system and help explain the performance difference, computational approaches are here applied to explore the molecular details in nanocarrier assembly process. By using synergistic computational and experimental approaches, microcosmic situations can be revealed and tracked and those details can help reveal reasons why and how subtle changes in building block affect micelles function. In this work, we conducted CG-MD simulations based on Luo's two latest represents telodendrimer structure $PEG^{5K}Rh_4CA_4$ and $PEG^{5k}CA_4$–LIN-Rh_4 and previous design $PEG^{5k}CA_8$[41-42]. The interactions among two critical building blocks and drug within atomistic and CG models: rhein (RHN), cholic acid (CHOA) and doxorubincin (DOX) are simulated and compared to validate CG model representation. Also, *in vitro* self-assembly processes were reproduced and tracked by using our simulations. Three generations of telodendrimer were self-assembled with doxorubicin in water solvent within 20% drug encapsulation rate and micelle morphologies were compared. Several explanations based on MD simulations results were given to the different performance among these three generations of telodendrimer.

In MD simulations, both atomistic model and CG models were used. Atomistic models were used to show the exact representations of each building block in molecules. CG models were derived from atomistic models, and they were used to simulate more complicated system for longer time. With the application of atomistic model and CG model, the multiscale MD simulations could show

chemical characteristics details of molecules as well as interactions among macro molecules. Our results showed consistency with real experimental results. And, our models were able to gives an insight for designing drug nanocarriers.

2.2 System Preparation

2.2.1 Coarse-grain model building

Coarse grain model of building blocks in $PEG_{5k}Rh_4CA_4$ and $PEG_{5k}CA_4$ –LIN-Rh_4 used in this thesis were kindly provided by Wenjuan Jiang shown in Fig 2. According to the valuated PEG_{5k}CA8 model reported in previous study, 4 cholic acids building blocks were replaced by rhein in PEG_{5k}Rh4CA4 to adapt Doxorubincin(DOX). Then, an additional link was added to $PEG_{5k}Rh_4CA_4$ structure and building blocks were reorganized to build third generation $PEG_{5k}CA_4$ –LIN-Rh_4 shown in Fig3. Each coarse-grained bead was assigned into four main types of interaction sites: polar (P), nonpolar (N), apolar (C), and charged (Q). According to Martini force field, each particle type was represented in more accurate representation in 18 subtypes to show chemical nature such as hydrophilicity and hydrophobicity.

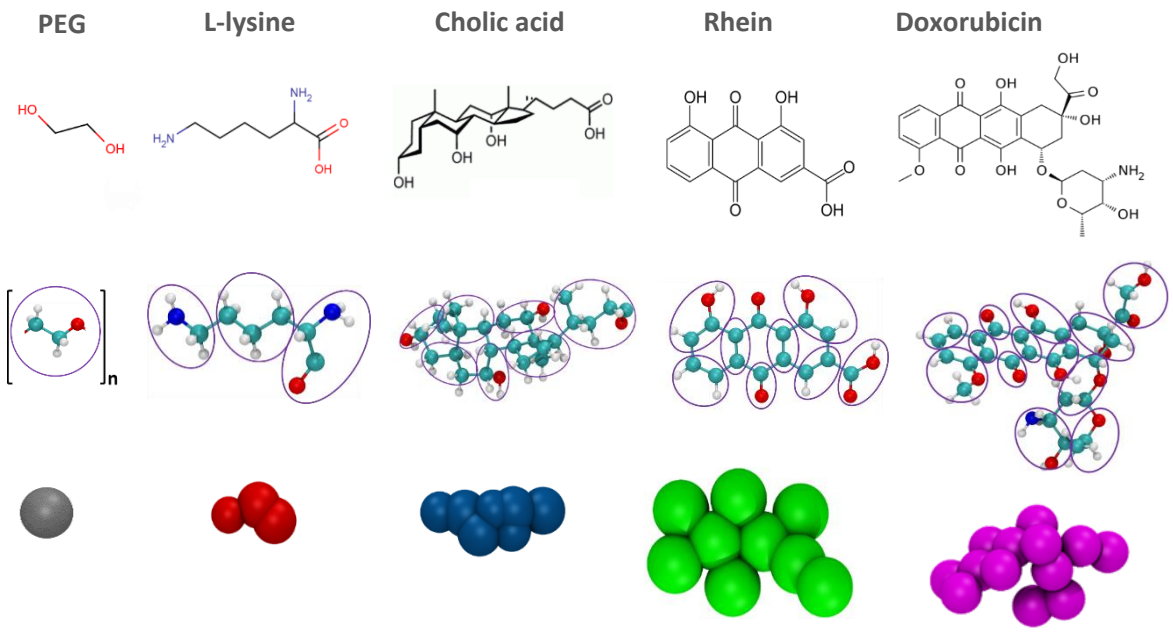


Figure 2 Coarse grain models of building blocks used in $PEG^{5k}CA_8$, $PEG^{5k}Rh_4CA_4$ and $PEG_{5k}CA_4$ –LIN-Rh_4

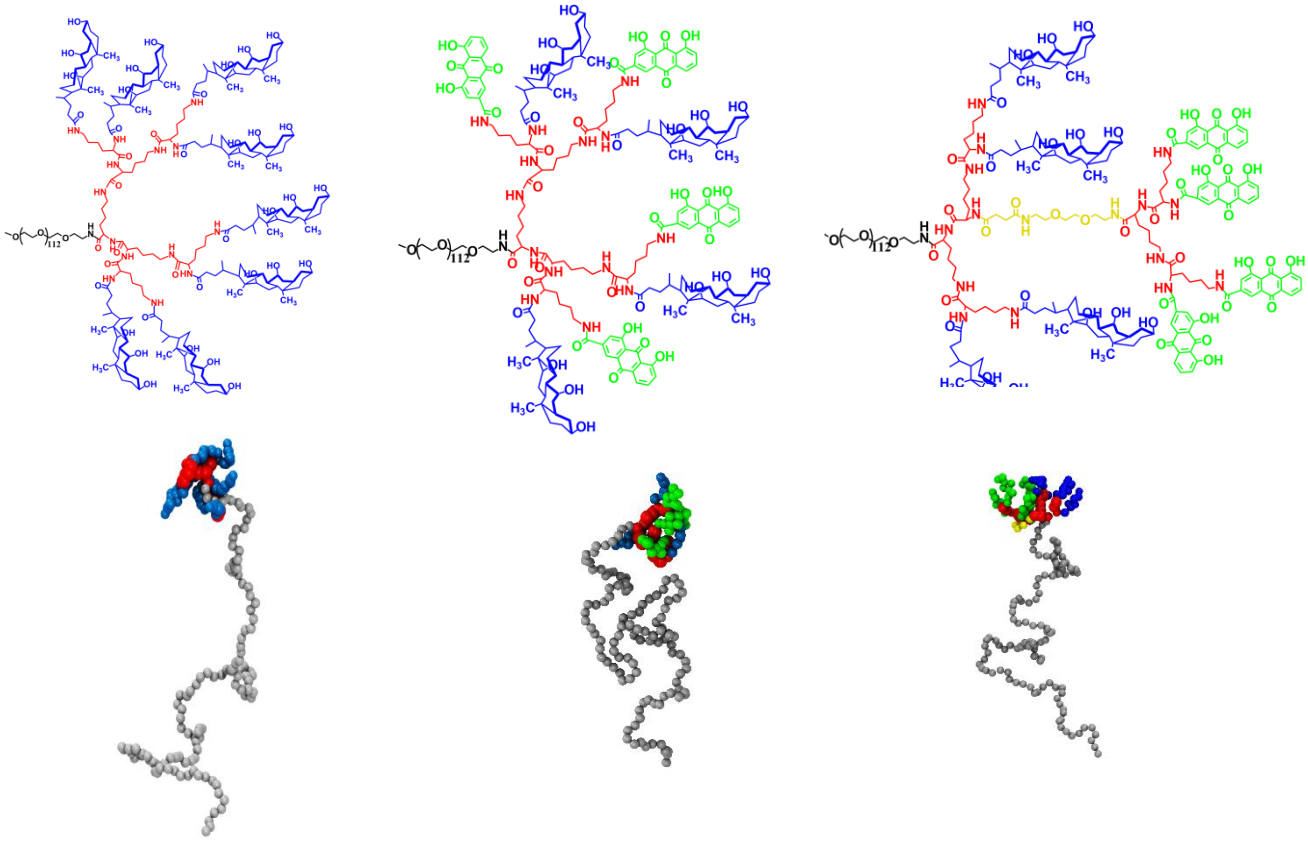


Figure 3 Coarse grain model of PEG5kCA8, PEG5kRh4CA4 and PEG5kCA4 –LIN-Rh4. Black parts: Polyethylene glycol (PEG); Red parts: Lysine; Blue parts: Cholic acid (CA); Green parts: Rhein (RH); Gold parts: link (LIN)

2.3 Simulation Details

2.3.1 Atomistic Benchmark Simulation

Molecular dynamics simulations on the coarse-grained model were performed with GROMACS 4.6.5 software package.

To validate CG model, benchmark simulations were performed to determine the difference between atomistic and CG models. As Table 1 shown, in both atomistic system and CG system, 18 of CHOA and RHN and DOX were randomly distributed in box filled with water initially. The reason only enabling CHOA and RHN rather than LYS and PEO interacted with DOX is that CHOA and RHN are the most critical drug attaching group with DOX, and their interaction performances would affect self-assembly directly.

In atomistic model, each building block was randomly packed in a 6nm^3 size box with filled of atomistic water molecules initially. The initial system was conducted for 0.05 ns energy minimization by 50,000 steps of 0.001 ps under steepest descent minimization, 0.02 ns NVT by 10,000 steps of 0.002 ps under Velocity rescale temperature coupling and 0.1 ns NPT by 50,000 steps of 0.002 ps under Velocity rescale temperature coupling and Parrinello-Rahman pressure coupling. All bonds (even heavy atom-H bonds) were constrained in the equilibration works in NVT and NPT process and temperature in each process was set at 300K. After well equilibrated, MD were run for 280ns with no position restrain.

In CG model, each building block was randomly packed in a 12nm^3 box with filled of CG model water. Standard MARTINI water involving 10% of anti-freeze water molecules (WF) was used here, which the WF beads are a little big than the normal CG water beads. The reason for adding antifreeze water is by using antifreeze water mole, will keep fluidity active and is more likely to real natural water molecules. The CG model systems were performed similar equlibration as

atomistic model while within larger time step and longer time. Energy minimization was performed for 2 ns by 50,000 steps of 0.040 ps under steepest descent minimization, 0.2 ns NVT by 10,000 steps of 0.020 ps under Velocity rescale temperature coupling and 1 ns NPT by 50,000 steps of 0.020 ps under velocity rescale temperature coupling and Parrinello-Rahman pressure coupling. After equilibration, CG models were conducted MD without position restraint for 1000ns. Both atomistic and CG models consider the periodical condition at xyz dimensions.

System	CHOA	RHN	DOX	SOL	Box size(nm)	Time
Atomistic	18	18	36	5557	6^3	280 ns
CG	18	18	36	16555W+1835WF	12^3	1000 ns

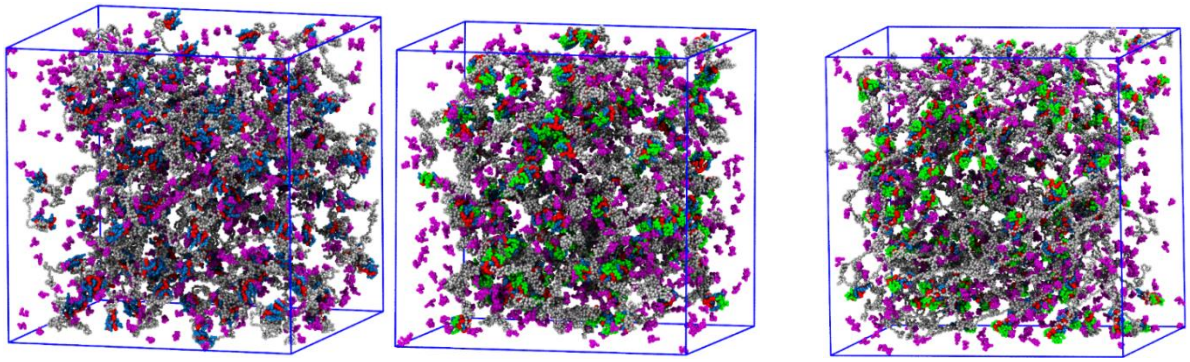
Table 1 Setup of benchmark system

2.3.2 Coarse Grain Self-Assemble Simulation

In order to replicate the self-assemble process as experimentation work, three CG molecular dynamic simulation systems were performed with each setup system shown in Table 2. In each system, 135 DOX and 614 Telo were randomly distributed in the system as Fig 4. The content ratio (ratio of drug loading) were around 23%, which is the ratio reported in Luo's researh. The content ratios in three system is also very close so that wouldn't affects our results due to ratio differnce even molecular weight of each generation telodendrimer was different. Then three systems were filled with standard MARTINI water molecules with 10% of antifreeze water molecule[23b].

System	DOX	Telo	W	WF	Drug Loading	Box size setup(nm)
$PEG^{5K}RH_4CA_4$			240679	27035	23%	
PEG^{5K}–CA_4-LIN-RH_4	135	614	287600	32970	23%	35
$PEG^{5K}CA_8$			251792	27538	22%	

Table 2 Setup of self-assembly system



(a) $PEG^{5K}CA_8$ (b) $PEG^{5K}RH_4CA_4$ (c) PEG^{5K}–CA_4-LIN-RH_4

Figure 4 Initial setup of three system. In order for clarify, solvent is hidden in these three images for clearness

Each system was first energy minimized by steepest decent algorithm for 2ns with 50,000 steps of 0.04 ps time step in order to adjust every components in the system to a natural geometries and orientation.

Two equilibration runs were performed after energy minimization.

An isothermal−isochoric NVT equilibration (number of particles (N), system pressure (P) and temperature (T) were conserved) run for 1 ns with 50,000 steps of 0.02 ps coupling the system at 300K based on energy minimized result structures. Temperature coupling was using velocity rescaling in a stochastic term.

An isothermal−isobaric NPT equilibration (number of particles (N), system pressure (P) and temperature (T) were conserved) based on NVT result structures, and run for 2 ns with 100,000 steps of 0.02 ps coupling at P = 1 atm and T = 300 K. Temperature and pressure coupling was using velocity rescaling stochastic term and Berendsen algorithm barostat term, respectively. Two groups were assigned, one group was water molecules (WAT in .mdp file) and another group was Telo and DOX (MIC in .mdp file), to be coupled in separately thermostats and isotopically barostat. The non-bonded interactions terms were computed with the standard cutoff of 1.2 nm. Using the standard shift function, the LJ potential was shifted from r = 0.9 nm to the cutoff distance so that both the energy and force smoothly vanish at the cutoff value.

After system was set up and equilibrated, CG model run molecular dynamics simulation for 3 μs at 0.02 ps step time. Coordinates, energy, and velocity trajectory files were saved at every 1, 0.1, and 0.2 nanosecond respectively.

2.3 Analysis Approach

2.3.1 Atomistic Benchmark Analysis

Pi-Pi stacking is a kind of non-covalent interactions between aromatic ring structures[43]. Due to the opposite charge interfaces on two different aromatic rings, pi bonds are formed to be stacks. Pi-Pi stacking is a very important interaction in micelle formation because it relates to the morphology of micelle and drug-nanocarrier interaction.

Because both DOX and RHN have aromatic ring structures. Pi-Pi stacking interaction occurred in three combinations of DOX+DOX, DOX+RHN, and RHN+RHN in the benchmark systems. They were visually seen and extracted from benchmark system results. Distance of Pi-Pi stackings were calculated by gromacs tool package *g_dist* and VMD 1.9.2 software distance

calculation. In CG models, since multiple heavy atoms were mapped to one bead, CG bead 2 on RHN and CG bead 2 on DOX which is close to center of plain were used to represent RHN and DOX plain. And the distances were calculated from Pi-Pi stacking configuration.

2.3.2 CG Self-assembly Analysis

In Self-assembly system, analysis in this sessions were using visual method.rames. Frames were extracted from the trajectory by using gromacs tool package *trajconv*. VMD 1.9.2 software were used track the trajectory of different groups.

2.4 Results and Discussion

2.4.1 Benchmark results

In order to apply CG model to conduct long simulation, we need to validate our CG model could provide same behavior and characteristics as atomistic model. Fig 5 shows details of benchmark system results. In snapshots, CG models show same behavior as atomistic model that pi-pi stacking structures were observed in DOX and RHN, RHN and RHN and DOX and DOX combination but not in DOX+CHOA combination in both atomistic and CG model. This is because DOX and RHN share planar ring structures with similar chemical characters whereas CHOA barely shows aromatic rings for pi-pi stacking. Given this difference, instead of forming stacking, CHOA are more likely to either attach to the surface of DOX cluster or simply cross-link to DOX cluster. And RHN and DOX can form pi-pi stacking thus contributing more to build micelle. On the other hands, it needs to be mentioned that multi-layer Pi-Pi stacking occurs in DOX-DOX interactions in both atomistic and CG models. Although it is not possible for multilayer structure for single aromatic ring structures because once pi-pi stacking form, charge on aromatic ring structure will be neutralized to prevent stacking grow, however, there are two aromatic rings in DOX structure that

both of them can be receptor for Pi-Pi stacking and multilayer pi-pi stacking can form by rotating DOX. Here, by properly building CG model, it can show same behavior pattern as the atomistic model.

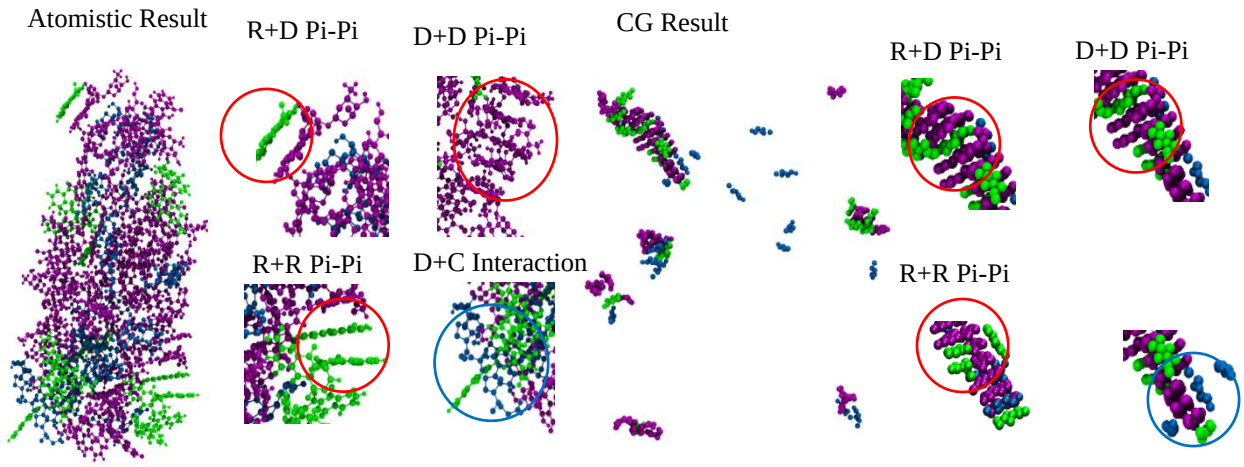


Figure 5 Details of RHN+DOX+CHOA benchmark results. Purple parts: DOX Black parts: Polyethylene glycol (PEG); Red parts: Lysine; Blue parts: Cholic acid (CA); Green parts: Rhein (RH); Gold parts: link (LIN)

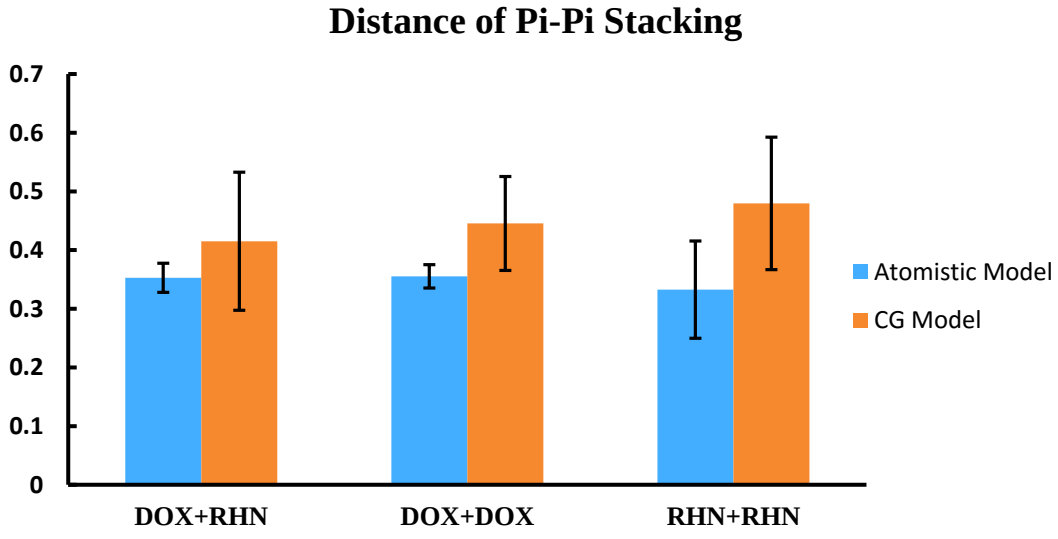


Figure 6 Distance of Pi-Pi stacking of three combination in atomistic and CG model

In order to further testify CG model, distance of Pi-Pi stacking in CG and atomistic model were measured and compared, the results are shown as Fig 6. The Pi-Pi stacking distances in atomistic for DOX+RHN, DOX+DOX and RHN+RHN were 0.353 ± 0.025 nm, 0.355±0.020 nm, and 0.333 ± 0.083 nm, respectively, while Pi-Pi stacking distances in CG model for DOX+RHN, DOX+DOX and RHN+RHN were 0.3415 ± 0.118 nm, 0.445±0.080 nm, and 0.480 ± 0.113 nm, respectively. The distance of pi-pi stacking in both models do not correspond to each other since an average of 4 heavy atoms are mapped to one bead that several coordinates are integrated together to make distance a little bigger. For that reason, CG model produces relatively longer distance of Pi-Pi stacking than atomistic model. Despite that, differences between corresponding CG and atomistic models keep constant, suggesting their similar performance in stacking which earns credit for CG model for modeling self-assemble process. It again proves that CG models' Pi-Pi stacking could provide same functions except for distance that wouldn't affect interaction.

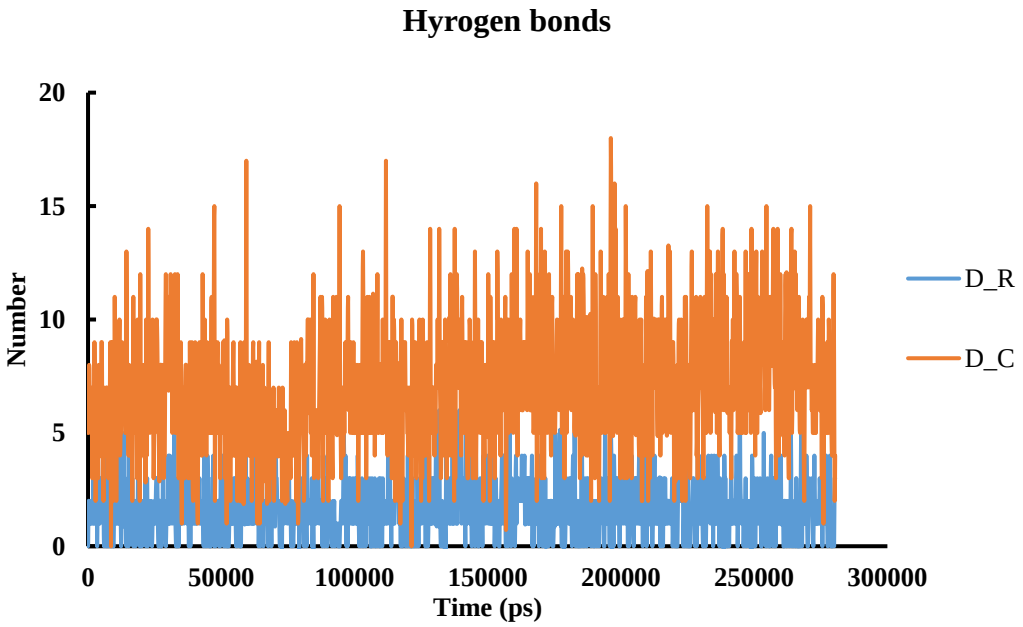


Figure 7 Number of hydrogen bond between DOX and RHN, and DOX and CHOA. Orange line: DOX and CHOA, Blue line: DOX and RHN

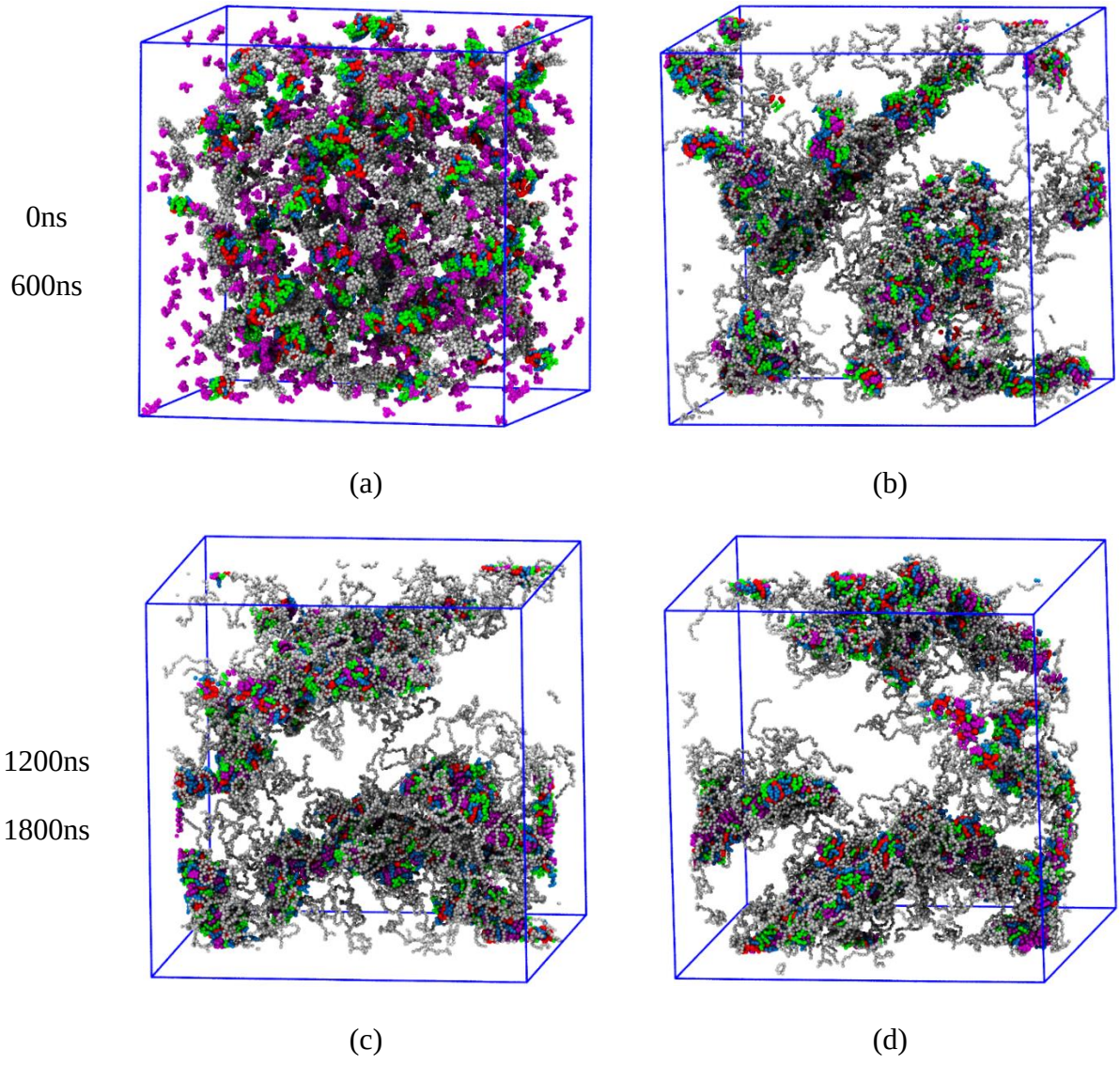
Hydrogen bond number between DOX+RHN, and DOX+CHOA in RHN+DOX+CHOA atomistic system were calculated and shown in Fig7. From the results, the number of hydrogen bond in both DOX+RHN and DOX+CHOA fluctuates within a small range which means both CHOA and RHN could interact stably with DOX. Specifically, D_C creates more hydrogen bonds than D_R, meaning that DOX and CHOA is more likely to interact by hydrogen bind rather than Pi-Pi stacking between DOX and RHN. This result also shows DOX interacting those two building blocks in different ways and preferences.

2.4.2 Self-assemble process of telodendrimers-DOX systems

In order to explore the different behavior patterns of telodendrimer-DOX interaction in $PEG^{5K}RH_4CA_4$, PEG^{5K}-CA_4-LIN-RH_4, PEG^{5K}-CA_8 systems and how different building blocks affect their self-assembly behavior, *in vitro* self-assembly processes were replicated as experimental works. Three systems were built within same number of telodendrimers and DOX whereas only difference among these three systems is the telodendrimer structures. All these three systems were randomly distributed initially and were set under same simulation environments.

In $PEG^{5K}RH_4CA_4$ system, the dynamic process of self-assembly is shown in Fig 8. From the serials of time-lapsed snapshots, DOX could be well encapsulated by telodendrimer and few DOX floats freely in the solvent at end of 3 microseconds. It has been clearly seen that, from 600ns, not only telodendrimer continued to capsulate DOX particles, but also the micelle pieces aggregate to form worm-like shape and later became a more complicated, branches structure. It drives attention that, the shape of micelle grows in several dimensions rather than simply elongating in one dimension. Massive branches were formed and physically crosslinked with each other to make the whole micelle thicker. At the end of 3 microseconds, three pieces of micelles were formed, that one big

branches worm-like shape and two short more spherical micelles. The simulation result matches the transmission electron microscopy image (TEM) in experimental data (Fig9e), showing the presence of both wormlike shape and spherical micelles.



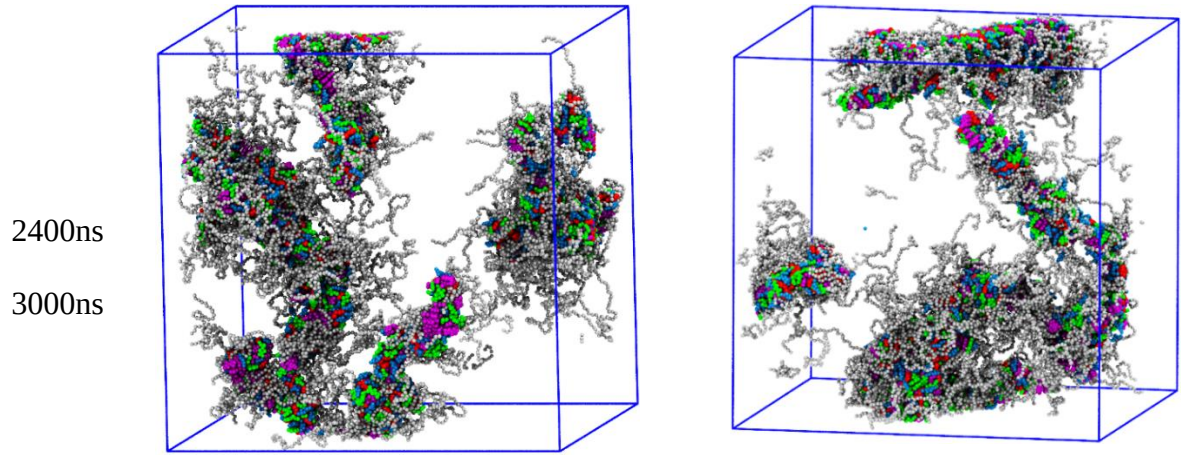
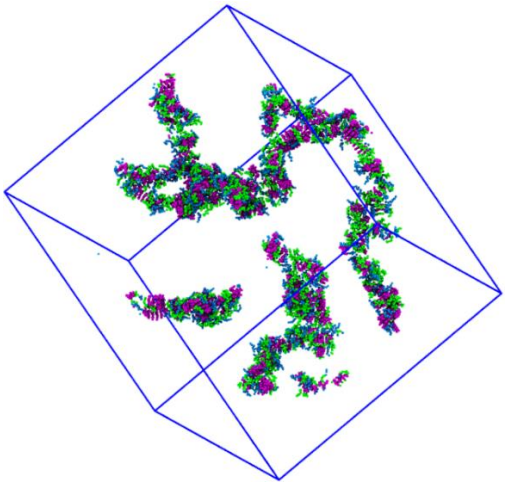
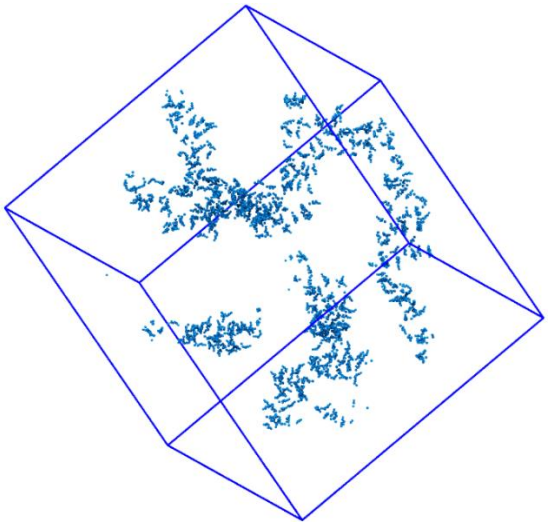


Figure 8 Trajectory of PEG^{5K}-CA_4RH_4 self-assembly process

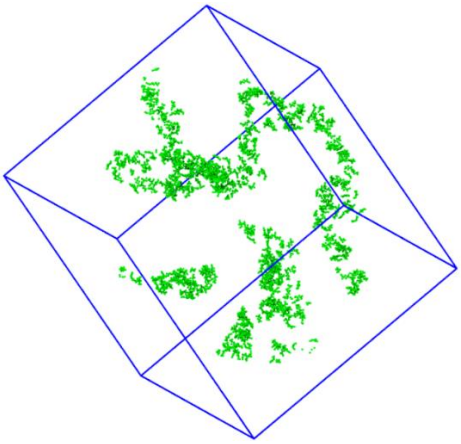
Figure 9 series images show the detailed construction of micelles in $PEG^{5K}RH_4CA_4$ system at 3 microseconds. By manually curtaining each group in system, the main core of micelle and aggregation situations can be clearly visualized and interpreted. The main core of micelle were built by DOX-DOX and DOX-RHN Pi-Pi stacking. As shown in figure 9d, up to 7 layers DOX Pi-Pi stacking was observed, constructing the domain of micelle core. Multi-layer Pi-Pi stacking makes the aggregation grow in one dimension, which is the main reason for the worm-like shape morphology. Another type of Pi-Pi stacking, T-shape Pi-Pi stacking which two aromatic ring structures are vertical to each other by favorable quadrupole, was observed frequently in Fig9. RHNs are forming T-shape stacking with DOX and help forming branches on the outside of DOX stacking micelle core as well as embedding in DOX stacking that help micelle core grow in one dimension. Moreover, from Fig9d, the branches were generally resulted from T-shape and sandwich Pi-Pi staking from DOX-DOX and DOX-RHN combination rather than crosslinking conditions from PEG chains.



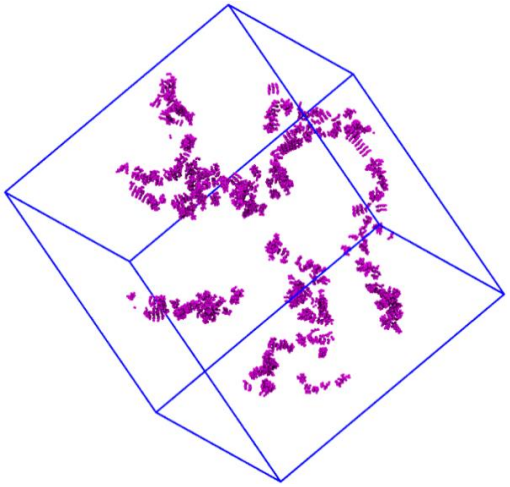
(a)



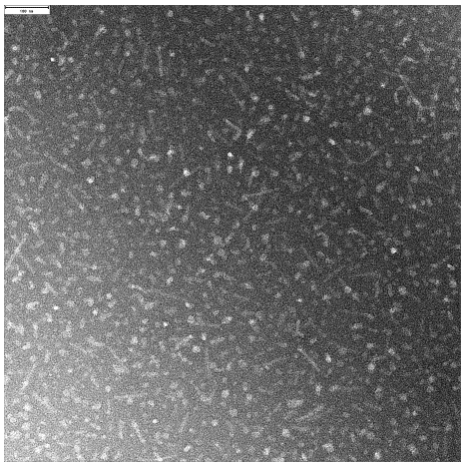
(b)



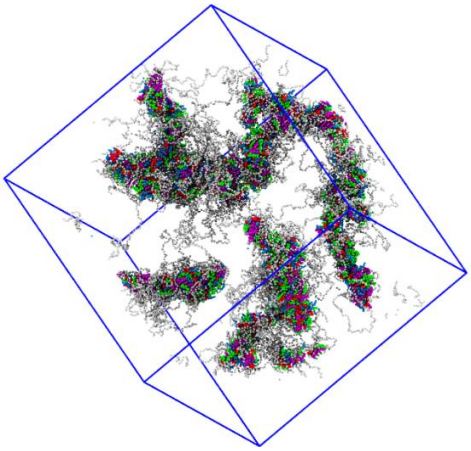
(c)



(d)



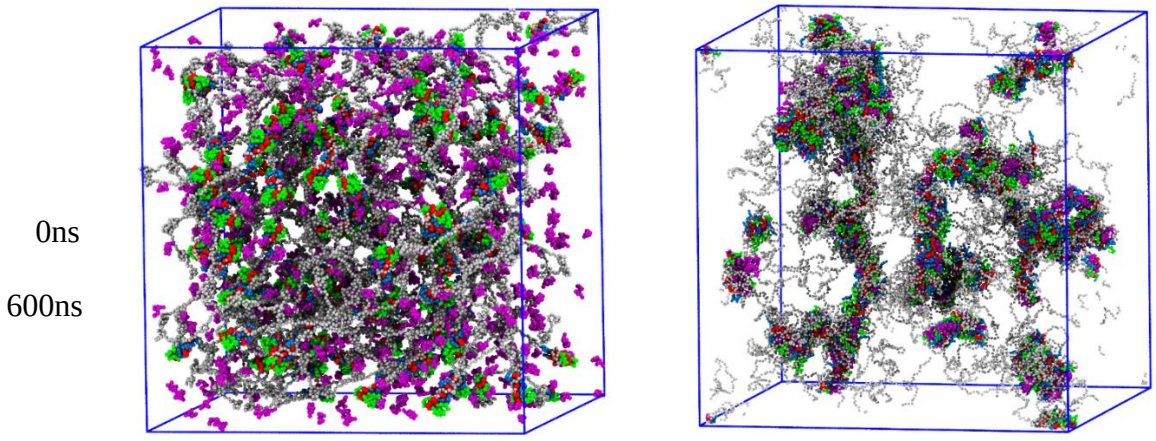
(e)



(f)

Figure 9 Snapshots of detailed construction of $PEG^{5K}RH_4CA_4$-DOX micelle, a. Micelle core without PEO; b. CHOA; c. RHN; d. DOX; e. TEM images; f. Whole micelle

In PEG^{5K}-CA_4-LIN-RH_4 system, the dynamic process of self-assembly is shown in Fig 10. At the beginning of trajectory, similar self-assembly behavior was found as PEG^{5K}-CA_4RH_4 system that micelle began to form. DOX was capsulated by PEG^{5K}-CA_4-LIN-RH_4 and few free DOX molecules floating in system. However, this system shows some difference comparing to the PEG^{5K}-CA_4RH_4 in each time frame. In serials of time frame, micelles aggregate more slowly than those in PEG^{5K}-CA_4RH_4 system, and DOX stacking are more likely to exposed to outside without covering by RHN. Besides, at end of simulation, the PEG^{5K}-CA_4-LIN-RH_4 system shows some linear micelle pieces without big branches were assembled rather than one big micelle with branches structure like PEG^{5K}-CA_4RH_4 system. Comparing the TEM image in experimental data (Fig11e), there is a good agreement that both wormlike shape and spherical micelles exist while there are wormlike shape micelles than that of in $PEG^{5K}RH_4CA_4$ system.



(a) (b)

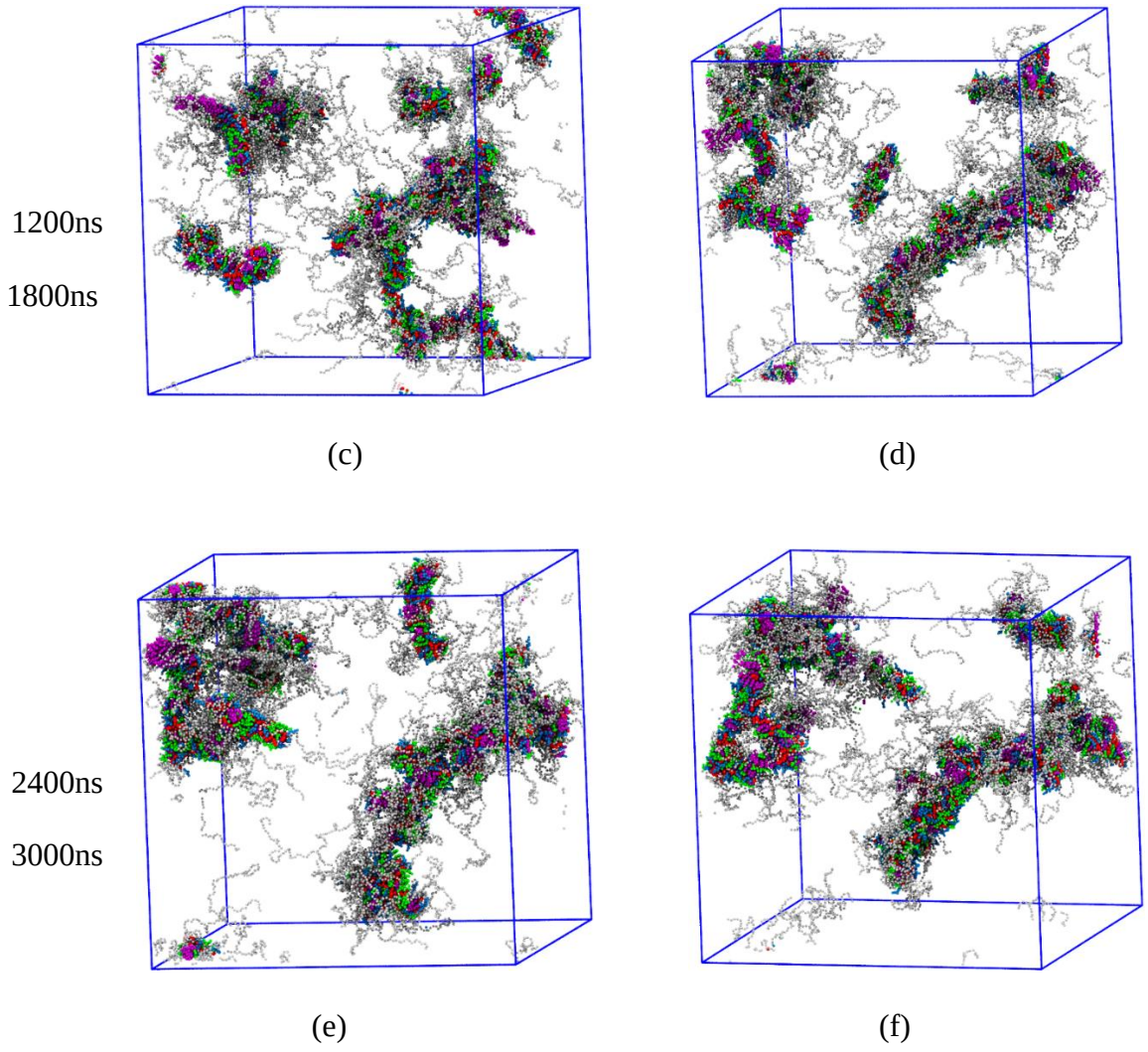
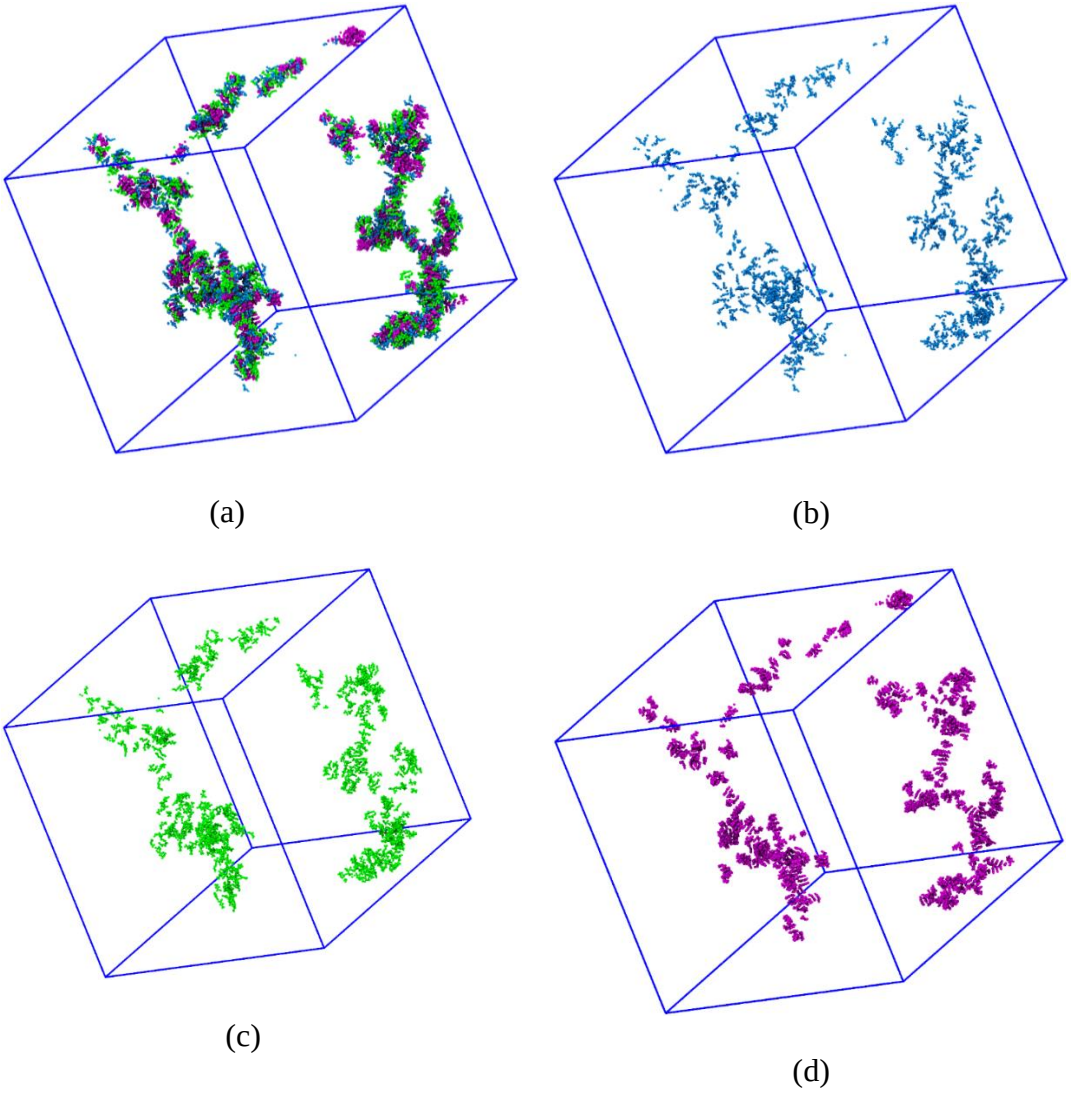


Figure 10 Trajectory of PEG^{5K}-CA$_4$-LIN-RH$_4$ self-assembly process

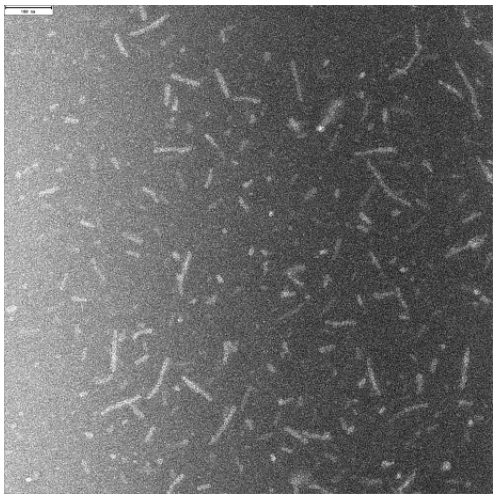
Figure 11 shows the detailed construction of micelle in PEG^{5K}-CA$_4$-LIN-RH$_4$ system at 3 microseconds. After hiding PEG chains, it is more obvious that there is less and smaller branches on the backbone comparing to the PEG^{5K}-CA$_4$RH$_4$ system. Even the main backbone of micelle were built by DOX-DOX and DOX-RHN Pi-Pi stacking like PEG^{5K}RH$_4$CA$_4$ system, it is also needed to be mentioned that DOX-DOX stacks were found being chopped by wider RHN-RHN stacks. No long multi-layer DOX Pi-Pi stacking like PEG^{5K}RH$_4$CA$_4$ system was found in PEG^{5K}-CA$_4$-LIN-RH$_4$ system. In another words, micelle core of PEG^{5K}-CA$_4$-LIN-RH$_4$ shows more

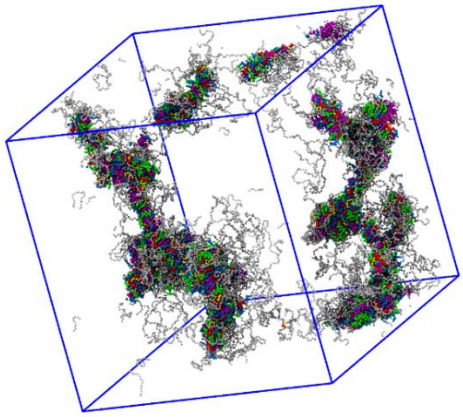
discontinuous than PEG^{5K}-CA_4RH_4 system. Instead of RHN forms T-shape Pi-Pi stacking with DOX, RHN will more likely to aggregate with RHN and insert to DOX stacks.



(a) (b)

(c) (d)





(e) (f)

Figure 11 Snapshots of PEG5K-CA4-LIN-RH4 telodendrimers at time t=3μs. a. Micelle core without PEO; b. Choa; c. RHN; d. DOX; e. TEM images; f. whole micelle

In PEG^{5K}-CA_8 system, the dynamic process of self-assembly is shown in Figure 12. It shows very different self-assemble behavior that DOX were poorly capsulated by Telodendrimer. Even spherical micelles were formed, the micelles were not stable at all during the simulations that the morphologies were keep changing. In spite of all DOX capsulated, PEG^{5K}-CA_8 system shows smallest spherical micelles among three systems. Micelles were very short and no clear wormlike shape was built. Also, in some micelles, micelle core DOX were unevenly exposed to outside environment. CHOA did attach to the DOX stacks by hydrogen bonds, but they were less likely to make any effort to form micelle and build branches. This result shows agreement with the experimental results that PEG^{5K}-CA_8 would precipitated within self-assembly with DOX.

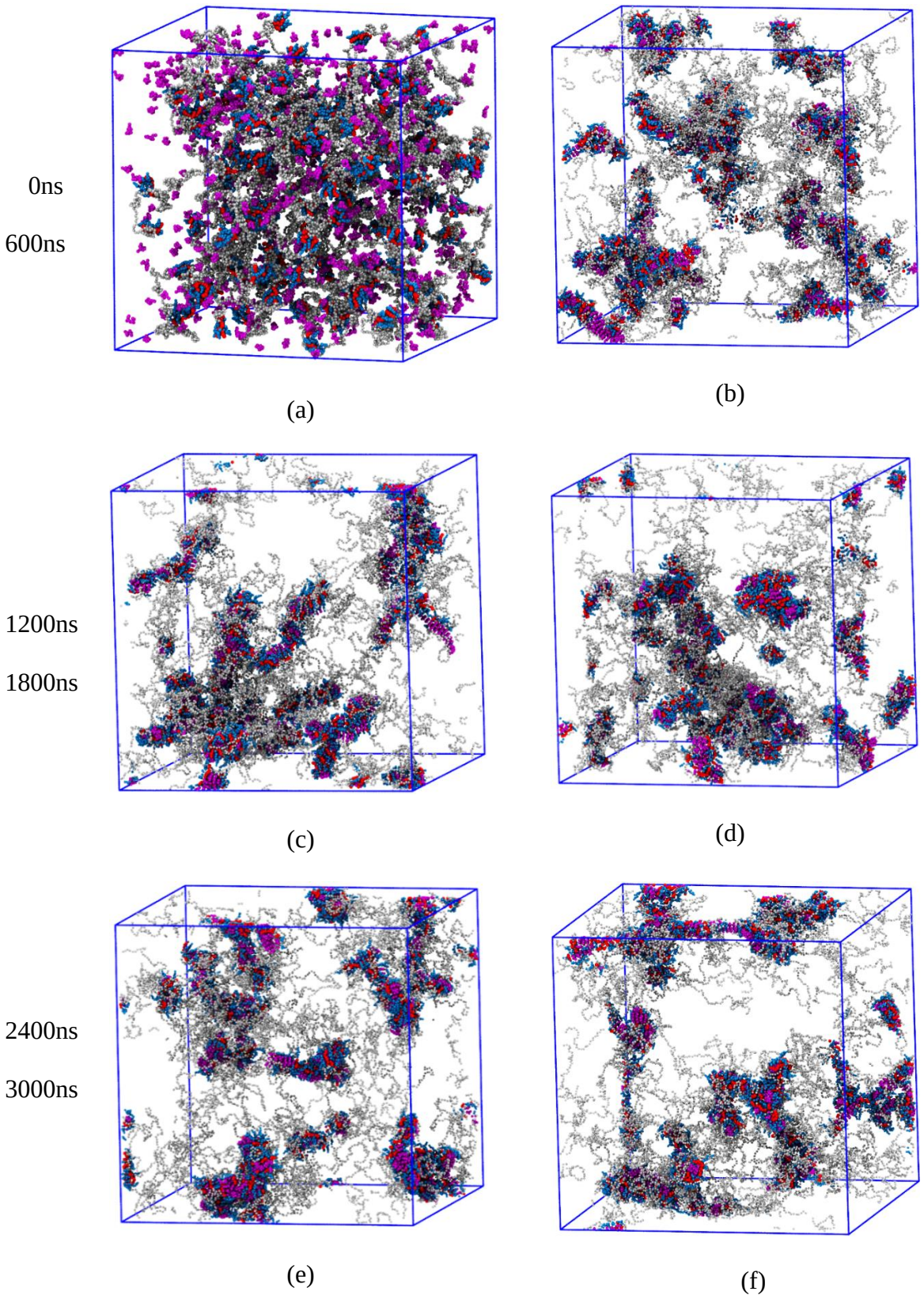


Figure 12. Trajectory of PEG^{5K}-CA_8 self-assembly process

Comparing these three systems mentioned above, some molecular and atomistic details which are not feasible in experiments are shown in computational way. Difference in micelle entire morphology and micelle core forming details were discussed above. According to the benchmark results, it is shown that RHN and CHOA share different interaction way with DOX. RHN is able to form Pi-Pi stacking with DOX which interaction is more stable than simple attaching driven by hydrogen bonds. Different interaction pattern result in different performance between PEG^{5K}-CA_8 and other two generations of RHN containing telodendrimer that PEG^{5K}-CA_8 would not form a stable micelle with DOX because CHOA would not help DOX micelle core grow, while two generations of RHN containing telodendrimer would. Difference in PEG^{5K}-CA_4RH_4 and PEG^{5K}-CA_4-LIN-RH_4 were also discussed that PEG^{5K}-CA_4RH_4 forms more big branches than the PEG^{5K}-CA_4-LIN-RH_4. This phenomenon could be explained by the chemical structure. Although both structures have same number of RHN and CHOA, their location in structure leads to the behavior difference. In $PEG^{5K}RH_4CA_4$, pairs of RHN and CHOA were located on LYS branches and all four RHN were separated by CHOA. It was mentioned previously that Pi-Pi stacking could occur in RHN-RHN combination. So, CHOA in the structure performs segmentation functions that prevents RHN forms Pi-Pi stacking with themselves. CHOA would not participate in stack, instead, it would control the number of RHN embedding into DOX-DOX Pi-Pi stacks. Therefore, paired RHN and CHOA can cover evenly over the DOX stacks which provides additional protection for drug from releasing from micelle. On the other hands, since RHN were separated and distributed evenly on the LYS branches, each of them has more chances to form Pi-Pi stacking with DOX that form branches in multiple directions on micelle core. When it comes to PEG^{5K}-CA_4-LIN-RH_4 system, different situation happened. A link chain was added into the whole structure and RHN and CHOA

were reorganized. The RHN-CHOA pair in on LYS branches was changed to two RHN-RHN pairs and two CHOA-CHOA pairs and four RHN and four CHOA were separated by linker. Without physical segmentation of CHOA, RHN pairs are more possible to form Pi-Pi stacking by themselves first rather than forming Pi-Pi stacking with DOX stacks. Then when the entire RHN-RHN stacks embedded into DOX stacks micelle core, they can only help micelle core grow in one dimension. With less help of RHN, branches are less observed in PEG^{5K}-CA_4-LIN-RH_4 system and micelles would grow longer in one dimension. This rational explanation could help interpret why longer micelles are form in PEG^{5K}-CA_4-LIN-RH_4 system. Besides, the longer linear micelle will lead to fragile micelle construction because it would bear more force in the middle point and more possible to break in middle, therefore leading to high drug release ratio.

2.5 Conclusion

Understanding nanoparticles' self-assemble process is crucial to determine characteristics of micelles, and MD simulation is able to provide some explicit details that benefits us for optimizing drug nanocarriers. In this thesis, long time scale molecular dynamics approach with coarse grain model was used to replicate real self-assemble process and match experimental data.

Three systems were simulated, explored and discussed in this project. Both PEG^{5K}-CA_4-LIN-RH_4 and PEG^{5K}-CA_4-LIN-RH_4 could form wormlike shape and spherical with DOX while PEG^{5K}-CA_8 could not. PEG^{5K}-CA_4-LIN-RH_4 system shows most complicated structure with branches, PEG^{5K}-CA_4-LIN-RH_4 system show more long micelle pieces with less branches. All these three system begin with random system with same number of Telodendrimer and DOX. Morphology results show consistency with Luo's research. It's clear that different construction of RHN groups in telodendrimer structure facilitate different interaction with DOX. Segmentation of RHN would

probably help micelle core grow in multiple direction and form more complicated micelle morphology.

3 Molecular architecture of tight junction proteins

3.1 Molecular architecture of tight junction proteins

Tight junctions (TJs), are multimolecular complexes constituted by intramembrane particles strands and fibrils associating the areas of two cells that regulate permeability of epithelia[44]. In 1972, Fromter and Diamond found whether the epithelia is leaky or barrier is dependent on the intercellular tissue at adjacent area of two cells which then so-called as "tight junction" (TJ)[45]. TJs can function as selective barriers, ion channels and fence based on the different protein complex compositions. Also, TJs have multiple functions such as stabilizing cells together, maintain the polarity of cells, preventing/transporting passage of molecules and ions, and etc[44a, 46]. With the understanding of function and architecture of TJs, challenges in intricate disease such as Alzheimer's[13], Parkinson's[14] and hepatitis C virus[12] are expected to more likely be solved.

Claudin, is a critical components in TJs, named from the Latin *claudere* means "close", first found and defined by Mikio Furuse and Shoichiro Tsukita in 1998[47]. Currently, within more than 40 kinds of proteins classified in TJs, such as occluding, actin, claudin, zona occludins, and junction adhesion molecules, claudins family transmembrane protein are proven to be vital compositions to determine the paracellular permeability in TJs[46, 50].With decades of study, 27 type of mammalian claudin has been found and classified which are expressed in all known epithelial tissues as epidermis, eye, cochlea, retinal pigment epithelium, exocrine pancreas and urinary bladder[48]. Claudin is a kind of small protein with molecular masses of 21-34 kDa, and its general structure contains four transmembrane domains (TM1-TM4), two extracellular loops (ECL1 and ECL2), an

intracellular NH2 terminus and an intracellular COOH terminus[49]. Claudins establish linear macromolecular strands with in TJ and provide a molecular seal to connect cells[47]. Existing experimental works have proven there are two modes *cis*- and *trans*- for claudin to interact and aggregate to be a macromolecular strands in TJs[49]. *Cis*- interaction means claudins are interacted within in the same membrane of a single cells and *trans*- interaction means claudins are interacted across two cells. Both two interaction modes are important for claudin to function as barriers or channels, but limited existing work could give explicit resolution of architecture-function relationship.

Currently, with the release of crystal structure of claudin-15 and claudin -19, structural study for claudin macromolecular strand began to flourish[51]. Through isolation of multi-order oligomers and mutation works, some potential models of macromolecular strands were predicted. There are many evidence from biochemical, molecular biology, cell biology area showing the claudins' functions are replied on their secondary and quaternary structure. For instances, Yu et al found the permeability of claudin-2 was based on ECL1 part, and mutating claudin-2 negative charged D65 on the ECL part will result in cation ion permeability decreased on MDCK I cells[46, 52]. Piontek J found ECL2 on claudin-5 plays an important role in trans- claudin-claudin interaction that aromatic residue F147, Y148, Y158 will perform a strong binding core[53]. Hiroshi Suzuki et al reported a dimer width strand model for homo claudin macromolecular strand, and it is constructed mainly by two dimer types which are mediated by ECL β-sheet domain and TM domain[54].

Although *cis*- and *trans*- interactions for types of claudins are studied and some intelligent models were built, those predicted models are based on simple homo- interaction that are not accurate to show the complexity of real claudin strand. They claimed claudin strand formations are mainly based on and resulted from one or two dimer types that made the models nearly uniform[11, 51, 55].

More complicated models constructed above two dimer types were not feasibly touched by experimentations, since it is too hard to isolate entire strand in TJs.

Here we use *in silico* molecular dynamic simulation method to study the architectures of macromolecular claudin strands. Two channel forming claudin-2, and -15 and two barrier forming claudin -1 and -15 was selected to be studied. Even those four types of claudin share highly conserved residues and similar structures[56], they located and functionalize differently in human body. We hypothesize that there are some relationships between strand structures and their functions.

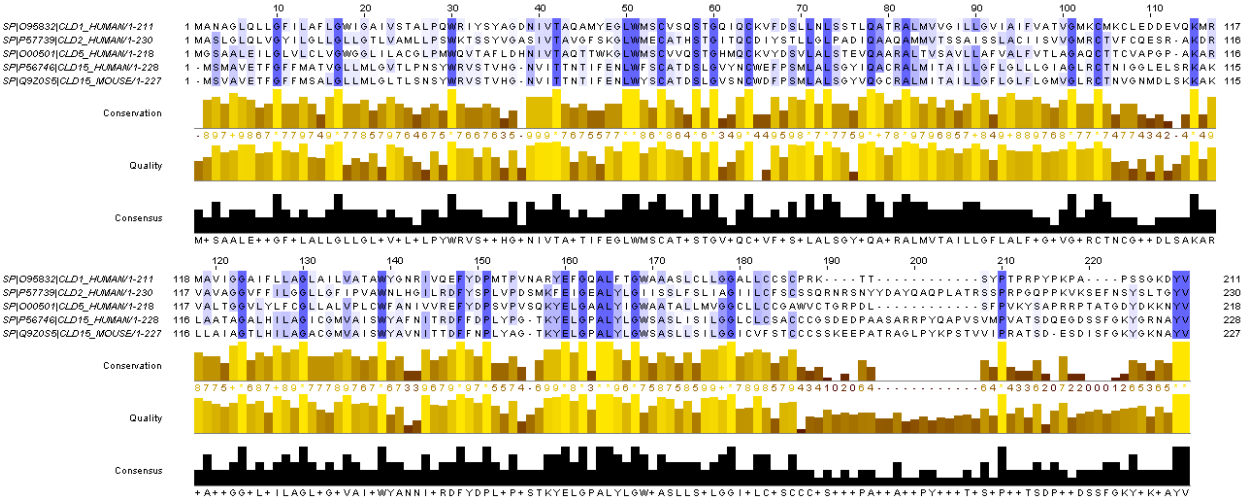


Figure 13 Sequence alignment of Human Claudin-1, -2, -5, -15, and Mouse Claudin-15. Sequence similarity is shown as blue.

A comprehensive study to construction of these claudin strand were built and process of claudin homo cis-interactions were simulated. In our simulations, claudin-1, -2, -15 were built carefully by homology modeling[15] according to the crystal structure of mammalian claudin-15[51]. Claudin-5 monomer structure is borrowed from lab member FJ Irudayanathan's work[57]. All three homology modeling results are validated by Ramachandran[16a] plot and Z-score[16b]. Four claudin structures testified by 100 ns atomistic level MD simulations, and they are proven to be stable and accurate

to applied long term MD simulations. But, the atomistic models are too computational expensive. CG models were introduced to improve simulation efficiency within acceptable accuracy. Self-assembly simulations are began with 64 monomer grid with random interfaces, and at the end of simulation, they aggregated into long strands which are comparable to freeze fracture images from experiments. Also, four representative dimer types were reproduced in our systems that show excellent agreements with models predicted by experiments[51, 55, 58]. Moreover, cluster analysis, dimer distribution and probability density map of each dimer type was performed to show the difference in strand formation processes among four claudins. Our results show classified Dimer A and Dimer C is predominately found in all systems that make most contribution to form the macromolecular strand backbone. We also find Dimer B are specially formed in barrier forming claudin, claudin-1 and claudin-5 rather than pore forming claudin, claudin -2, -15. It is also showed only Dimer B decreased after two amino acid mutation work, which can be seen as a potential dimer type in claudin-5 strand formations.

3.2 System Preparation

3.2.1 Homology Modeling

Homology modeling is a technique of constructing 3D atom resolution protein structure of targeting sequence according to the empirically determined structure of homologues protein. Since amino acid sequence determine the secondary structure, by align sequence with homologous protein, targeting protein structural model could be produced.

Claudin-5 structure was borrowed from FJ Irudayanathan's previous work[57] which has been properly prepared in homology modeling. The sequence of claudin-1, -2, -15 were downloaded from Protein Data Bank CLD1_Human (PDB ID: O95832), CLD2_Human(PDB ID: P57739), CLD15_Human(PDB ID: P56746) in the Protein Data Bank. Mutated claudin-5 sequence was manually revised based on claudin-5 sequence (PDB IDO00501).

Several severs were used for homology modeling. I-TASSER[59] (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and PHYRE2[60] (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) server was used to produce potential structure of claudin-1, -2, -15. In order to further refine claudin structure, four highest score structures of each claudin from I-TASSER and PHYRE2 server as well as crystal structure of mouse claudin-15 (PDB ID: 4P79) and Claudin-5 from FJ Irudayanathan's work were used as templates in homology modeling by Yasara Software. The structures from homology modeling were scored based on structure quality factor and membrane protein quality by using PROSA sever (https://prosa.services.came.sbg.ac.at/prosa.php) [61] and RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) server. Those low score parts in structures were built by QUARK[62] (http://zhanglab.ccmb.med.umich.edu/QUARK/) server and iteratively used as template to refine the structure.

3.2 Simulation details

3.2.1 Atomistic Simulation

Molecular dynamics simulations on the coarse-grained model were performed with GROMACS 5.0 software package[63].

Qualified atomistic structure of claudin-1, -2, -5, and -15 monomers were oriented and hydrophobic thicknesses were calculated in PPM server (http://opm.phar.umich.edu/server.php).

Molecular dynamics simulations on the atomistic model were based on CHARMM 36 force field. One claudin-1, -2, -15 monomer was embedded into POPC lipid molecules and surrounded by atomistic water and 0.15M KCL by using charmm sever (http://charmm-gui.org/?doc=input/membrane). The systems were in periodical conditions. The monomer of each claudin was first energy minimized by steepest decent algorithm for 5000 steps, then serially equilibrated by NVT and NPT. After systems were well equilibrated, MD was performed in atomistic model for 1 microsecond with 50,000,000 steps of 0.02 ps coupling by at pressure 1 atm and temperature 310.15 K. Temperature and pressure coupling was based on Nose-Hoover algorithm and Parrinello-Rahman, respectively. Verlet cut-off scheme was used in all minimization and equilibration for it is faster and more accurate compared to group scheme.

System	Protein	Number	Lipid Membrane	Time
AA	Claudin-1	1	POPC	100 ns
AA	Claudin-2	1	POPC	100 ns
AA	Claudin-15	1	POPC	100 ns

Table 3 System setup for atomistic simulation

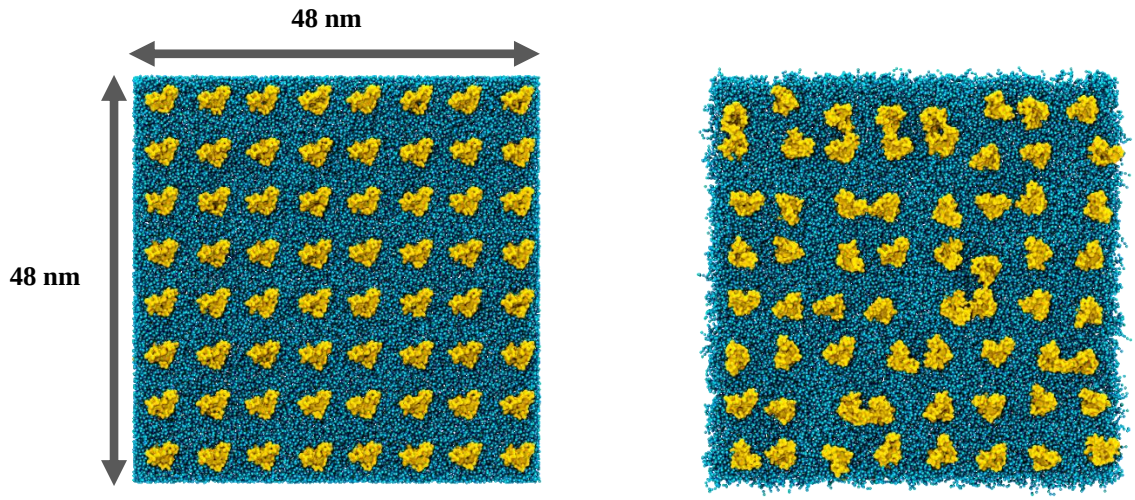
3.2.2 Coarse Grain Self-Assemble Simulation

Average structure of each claudin monomer was extracted from trajectory. Since atomistic monomer's shape was structurally changing during the MD process, structure from last frame will not be able to represent accurate monomer structure. Given the fact that C-terminal part cannot significantly affect claudin dimerization and strand formation except for diffusion coefficient due to lower overall mass, C-terminal residues beyond 195 were removed from the structures in order to reduce the box size in z-direction thus reducing the whole system volume and save computational efficiency. On the other hand, though C-terminal of laudin-1, -2, -5 and -15 starts at 185, 184, 181, and 183, respectively, the rest of C-terminals (~194) were not observed to have significant interaction during the MD simulations.

In order to build a 8×8 monomer grid, averaged atomistic monomers were initially placed in a 4×4 grid with inter center-of-mass distance of 6 nm using YASARA model software. Coarse grain model of each monomer in grid was built by MARTINI parameter set and its secondary structures was following ElNeDyn network. By using ElNeDyn network combined with coarse grained molecular force field, elasticity of protein scaffold can be well represented in molecular dynamics simulations. Insertion of protein in CG DOPC lipid bilayer was built by INSANE script. System was filled by MARTINI standard water molecules with 0.15M NaCl. Then the 4×4 grid was energy minimized for 0.5 ns with steepest decent algorithm and equilibrated by NVT and NPT for 100ns and 1000ns, respectively. Three groups were assigned in the index file. First group was water solvent and ions ("SOL" in .mdp file), second group is lipid ("LIP" in .mdp file) and third group was protein ("Protein" in .mdp file), to be coupled in separately thermostats and isotopically barostat. In NVT equilibration process, each groups were kept at 310K constantly with velocity rescale algorithm. In NPT equilibration process, semiisotropic pressure coupling was applied by

using parrinello-rahman algorithm, with a pressure of 1 atm. NVT and NPT process constrain all xyz dimensional position of each monomer.

8×8 grid was built from 4×4 grid as unit by using the same approach to make 4×4 grid from monomer. 4×4 monomer grid was duplicated, moved and organized to generate an 8×8 monomer grid with 6nm inter center-of-mass distance. 8×8 grid systems only ran 0.5 ns of EM, 50 ns of NVT and 20ns of NPT for energy minimization and equilibration because all 4×4 monomer grid unit has been well prepared. In order to avoid configuration bias in dimerization, the 8×8 grid systems ran additional NPT process for 200 ns with position restrain on 87GLY backbone bead and z axis rather than entire position constrain to make random orientation on xy dimension. The reason to choose 87GLY backbone bead to be position restrained is that it is inner bead close to center axis of protein, thus relative position among each monomers would not be changed except for the orientation on xy dimension (Fig14).



(a) (b)

Figure 14 Claudin-1 8×8 monomer grid. (a) 64 monomers within same orientations (b) represents all monomers with random orientations.

After systems were set up and equilibrated, 8×8 grid system underwent molecular dynamics

simulation without positional restraints at 20 fs step time. Coordinates, energy, and velocity trajectory files were saved at every 0.5 nanosecond.

64 monomer system setup was shown in Table 4. Each system was repeated under the same simulation environment.

System	Protein	Number	Lipid Membrane	Time*Repeat Times
CG	Claudin-1	64	DOPC	3 μs*2+5μs*1
CG	Claudin-2	64	DOPC	3 μs*2+5μs*1
CG	Claudin-5	64	DOPC	3 μs*9+5μs*1
CG	Claudin-15	64	DOPC	3 μs*2+5μs*1
CG	Mutated Claudin-5	64	DOPC	3 μs*10
			Total	97 μs

Table 4 System setup for CG simulation

3.3 Analysis Approach

3.3.1 Reverse Mapping

The coarse grained system from MARTINI were reverse mapped to atomistic system in CHARMM36 by using initram.sh script published by Wassenaar et al[64]. The reverse mapping process includes projection, energy minimization and serial position restrained NVT simulations. In our work, all CG models were energy minimized for above 0.2 ns and short MD of 0.02 ns with atomistic topologies. Proteins group was extracted from the system and monomers were reverse mapped for analysis. Solvent and lipids were hidden here for we focus on dimerization configuration here,

3.3.2 Monomer Analysis

Structures resulted from Homology modeling were scored by structure predict server mentioned in *section 3.2.1*. Ramachandran plot and Z-score were used to validate the structure of claudin from homology modeling results. After short time atomistic simulation(mentioned in *section 3.2.1*), atomistic monomer structures were evaluated by root mean square deviation (RMSD), root-mean-square fluctuation (RMSF), number of hydrogen bond (g_hbond) package tools in Gromacs. Donor-acceptor cut-off distance used in calculating H-Bonds were set as 3 Å.

Root mean square deviation calculates the overall configurational change of protein structure according to the initial position during the time. Root mean square fluctuation calculates the flexibility of atomic positions in the trajectory during the time. Hydrogen bond is critical element to maintain the secondary structure, number of hydrogen bond can reflect the stability of secondary structure.

3.3.3 Distribution of Dimer Type

Counting dimer type approach was based on R programming which is a software widely used in bio-information area. A script to determine the dimer type was developed to count the each dimer type and details are attached in Appendix.

Before counting the dimer type, four configurations of dimer were found and defined by published work[51, 55b, 65]. '*' symbol is used to clarify a second monomer. Dimer A is formed from residues on ECL1 and ECL2*-TM3*-TM4* by opposite charges and hydrophobic-hydrophobic interaction. Dimer B is mainly mediated by TM3-TM3* interaction. Dimer C is mediated by residues on ECL2, TM3 and TM4 interaction. Dimer D is formed by an anti-parallel β-sheet arrangement of ECL-1 residues. Configurations are shown as Fig15.

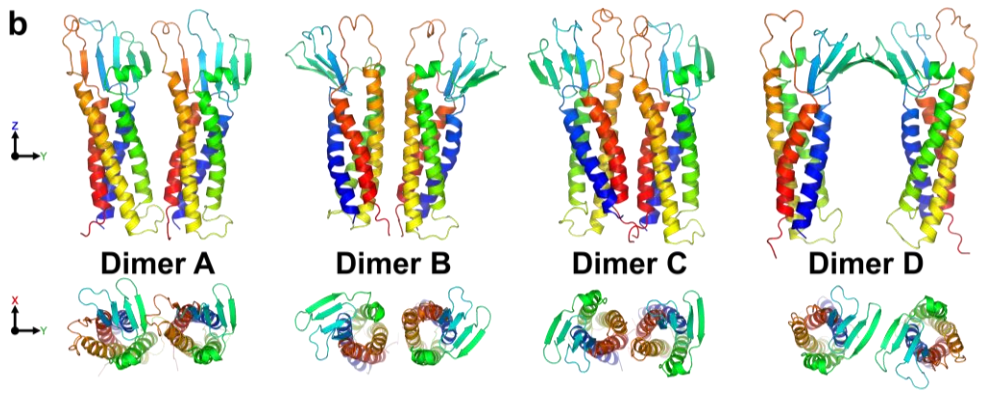


Figure 15. Configuration of Dimer. Blue segment TM1, green segment TM2, yellow segment TM3, red segment TM4.

The script for determining these 4 types of dimer was built based on these models' configuration. Several residues were picked to represent each TM part of claudin monomer as table 3 shown. Some residues were specifically picked according to the published work as criteria to specify dimer type.

Criteria to select residue would be (1) They are conserved residue in all four claudins (2) They are located in the middle of each TM to avoid distance bias from tilt (3) They have similar in structural and chemical property. (4) Backbone Cα atom of each residue was picked

	TM1		TM2		TM3		TM4	
Cld1	8-28	Leu16	82-102	Leu90	116-136	Leu127	164-184	Leu174
Cld2	8-28	Leu16	82-102	Ile90	116-136	Ile126	163-183	Leu172
Cld5	8-28	Leu16	82-102	Leu90	116-136	Leu126	160-180	Leu173
Cld15	2-24	Leu17	75-99	Leu88	116-136	Leu126	160-182	Leu170

Table 5 Topology and selected represent residues for each TM domain of claudin

Dimer type is determined by the configuration of dimer models. Series of distance matrixes were

generated that TM-TM* distances and specific residue-residue* distances were calculated.

Those which distance is below cutoff will be classified as “interacted” and above cutoff will be classified as “apart”. Cutoff parameterizations were built by more than five samples in each dimer configuration which are not shown in this thesis. Detailed cutoff parameterizations are shown in Appendix.

Each dimer type is determined by following rules:

Dimer A: TM1 and TM2 interface interacted with TM2* and TM3* interface, ECL1 interacted with part of TM4*; TM3 and TM4 interface is apart from TM3* and TM4* interface; Orientation of monomer and another monomer’s should be similar (standard deviation of (TM1-TM1*) and (TM2-TM2*) is lower than 10)

Dimer B: TM3 interacted with TM3*, 138 TRP and 138 TRP* are interacted; TM1 is apart from TM1*, 147 PHE is apart from 147PHE*

Dimer C: TM3 and TM4 interface interacted with TM3* and TM4* interface, 180 CYS and 180* CYS are interacted; TM1 is apart from TM1*, 138 TRP and 138TRP* are interacted, 147 PHE and 147 PHE are interacted.

Dimer D: Residue 63-65 on β-sheet interacted with residue 63*-65* β-sheet by anti-parallel arrangement, 4^{th} β-sheet is paralleled to 4^{th} β-sheet*

The TM-TM distance cutoffs for determining “interacted” were ranged from 0.2nm-0.24nm. The residue-residue distance cutoffs for determining “interacted” were ranged below 0.15nm. The cutoffs varies depend on the tilt conditions of claudin in lipid membrane. Those distances for

dimers tilting will have slightly more distance than those of dimers are vertical to lipid.

3.3.4 Conformational Analysis

Internal and external angle of dimer was calculated by R script and used to help explain dimer configuration. A script to determine the dimer Internal and external angle was developed by R to count the each dimer type and details are attached in Appendix.

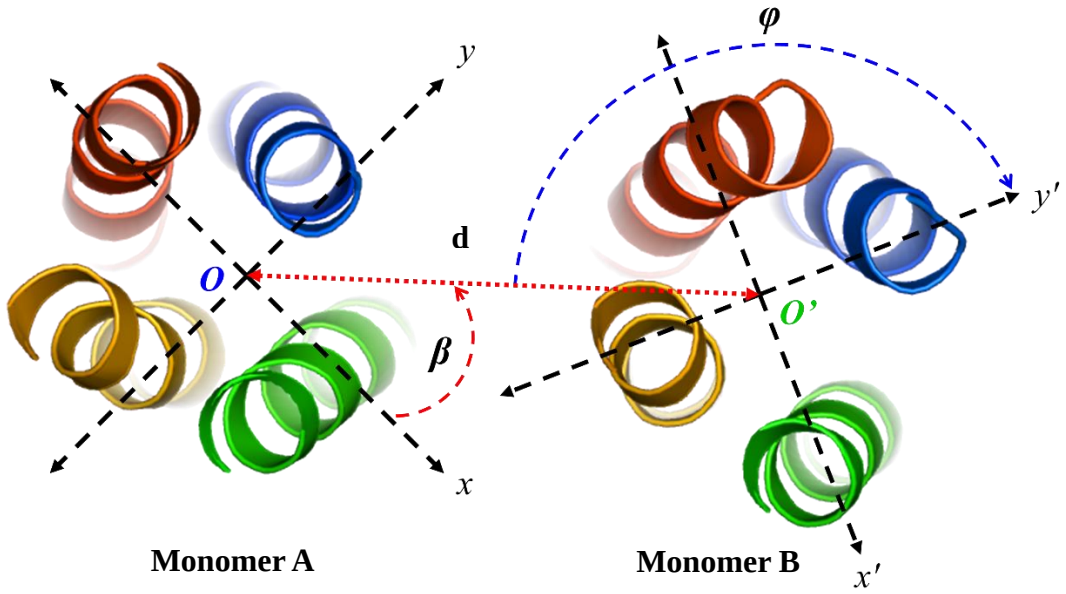


Figure 16 Representation of angle calculated in dimer

Firstly, centers of mass of all monomers were calculated to represent location monomer. All monomers which center of mass distances between two monomers below 3.0nm are regarded as dimer. Secondly, four points within the dimer were selected in dimer for angle calculation, center of mass of TM1 and overall monomer of two monomers. Coordinates system is built upon center-TM1 vector as positive y axis. Angle β and angle φ are calculated as Fig 16 shown. Angle β represents the external angle measuring the rotation of monomer B to monomer A and angle ϕ represents the internal angle of monomer B itself according to monomer A. The range of β and ϕ

is (-180^{o} , 180^{o}) and (0^{o} , 360^{o}). Thirdly, angle β and angle ϕ plots are performed and visualized by kernel density estimation map to estimate the probability density function of angle distribution.

3.3.5 Cluster analysis

Cluster analysis was used to track cluster change during the trajectory at 20 ns intervals (50 frames) by gromacs tool package *g_clustsize*. Number of cluster and max cluster size was calculated at the cutoff distance at 0.5 nm. All monomers within distance 0.5 nm will be considered as a cluster.

3.4 Visualization

Visualization of the simulations trajectories were performed by Visual Molecular Dynamics Software 1.9.2 and YASARA model software suites.

3.5 Results and Discussion:

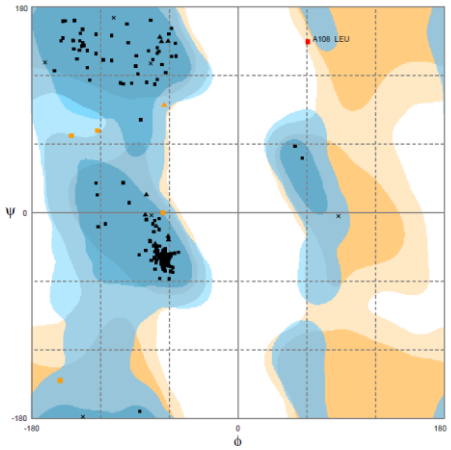
3.5.1 Homology modeling of Claudin-1, -2, -15

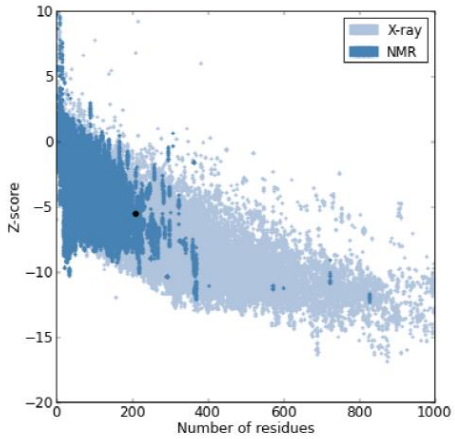
Human claudin-1, -2, -5 and -15 3D structure has not been solved and explored empirically by existing research. So, before conducting molecular dynamics simulation, we need to build monomeric structure by homology modeling based on homologues protein crystal structure. Here, we generated monomeric 3D structure of human claudin-1, -2, -15 based on the structure of murine claudin-15(PDB ID: 4P70) and previous work on claudin-5. The C-terminal domain and missing residues were modeled using *ab initio* method.

After serials of refinement and remodeling, the final structure qualities were evaluated by ramachandran plot and Z-score in several severs, and results are shown in Fig 17. Both Ramachandran plot map and Z-score of protein structure quality are vital step to testify predicted protein structure. Ramachandran plot map considers the energy within the protein structures and shows the rational areas of backbone dihedral angles ψ against φ of amino acid in all existing protein structures. Claudin-1, -2, -15 shows 99.5%, 100%, 99.1% region agreed with the ramachandran plot respectively which means secondary structure within protein are overall energetically reasonable. The LEU108 in claudin-1, ARG197 and MET206 in claudin-15 was in outlier region. However, they wouldn't take too much influence in our simulations because LEU108 is non-conserved amino acid in intracellular domain that make little effort to cis-interaction and ARG197 and MET206 are the amino acid domains above 194 which were removed from structure will be discussed afterwards. Z-score map shows the qualities of overall structure of all experimentally determined protein structures in protein data bank. Overall structure Z-score was calculated by X-ray and NMR, and our Z-score for each claudin was in the either X-ray area or NMR area which means our homology modeling structures can be typically found for native

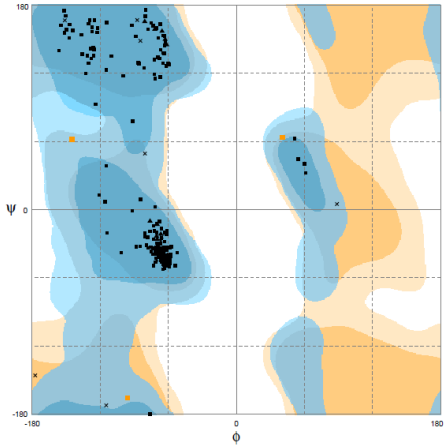
proteins of similar size. Claudin-5 model used here was borrowed from FJ Irudayanathan's previous work which 99% agreement with Ramachandran plot and the predicted secondary structure.

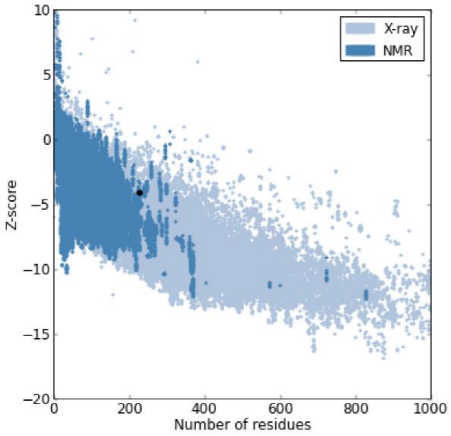
Claudin-1





Claudin-2





Claudin-15

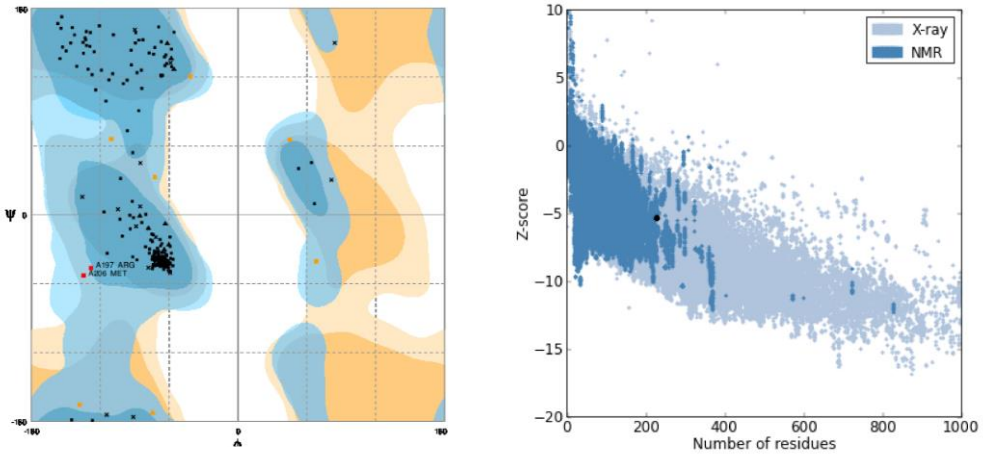


Figure 17 Ramachandran plot and Z-score of Claudin-1, -2, -15.

After all these three atomistic claudin structure from homology modeling were validated, they were inserted in DPPC lipid membrane and run in molecular dynamics simulations for 100ns without positional restraints. Root mean squared deviation (RMSD) of protein, root mean squared fluctuation (RMSF) and number of hydrogen bond were used to evaluate monomer stability. RMSD is a measurement of overall structural deviation comparing to the initial structure during whole trajectory, which can be seen as an indicator of conformation stability. If RMSD keeps fluctuating in a small range, then we can regard the conformation will not change too much and keep stable during a period of time. Instead of determine the whole structure stability, RMSF is a measurement of deviation of position among the atoms in a period of time which can reflect the flexibility region in the structure. The difference between RMSD and RMSF is that RMSD calculates average position change of overall proteins conformation over time, while RMSF calculates average position change of each atom/residue over time. Number of hydrogen bond is a critical measurement for testify stability of the secondary structure in the protein. Within a stable conformation, number of hydrogen bond should keep stable during the time, otherwise, fluctuation in number of hydrogen bonds means secondary structure is changing and unstable.

In Fig17, orange line in RMSD of claudin-1 shows RMSD for 194 amino acid domain claudin-1

model, which C-terminal domain was cut off, kept at a good convergence around 3.5 Å with no obvious fluctuation. Comparing with full 211 amino acid domain claudin-1 model shown in blue line that have a 2 Å range fluctuating RMSD around 5 Å, 194 amino acid domain claudin-1 model has a more stable structure during the whole time period. The fluctuation in RMSD was supposed to be resulted from the C-terminal domain. This inference was again proved by RMSF results. From Fig17 RMSF of claudin-1, there were peak fluctuations at 32-47 residue around, 101-119 residue around4 Å 7 Å and 190-211 residue around 11 Å, showing the part of ECL1 region, ICL and C-terminal part are highly flexible part in the simulation. Flexibility of C-terminal was speculated to result from the charged residue and phosphorylation sites. Although C-terminal of claudin make effort on trafficking to tight junction, protein degradation, PDZ-binding, and phosphorylation, there is no evidence on C-terminal will make effort on cis-interaction[49, 66]. Besides, removing C-terminal domain could help us reduce whole system volume on the z axis that improve the computational efficiency. From Fig17c, claudin-1 monomer forms 161±5 hydrogen bonds with no significant fluctuation during 100ns, which means that the whole secondary structure was stable in this period of time. For those results mentioned above, it is reasonable to claim that 194 amino acid domain of claudin-1 is a well prepared and ideal atomistic monomer structure for the following molecular dynamics simulations.

Same analysis approach was used to testify claudin-2 and -15. The results are shown in Fig17. With removal of amino acid above 195, claudin-2 RMSD was decreased to a lower level around 4.5 Å and there is no significant fluctuation. Highest flexible regions were in ECL1, ICL, and C-terminal domain, which is similar with claudin-1. The hydrogen bonds in claudin-2 was kept at 160±7 during 100 ns. Claudin-15 194 amino acid domain RMSD was kept stable around 3.4 Å. Claudin-15 RMSF showed two peak in ECL1 domain aroung 37, and 56. These two obvious

flexible regions were found to consistent with published crystal structure of mouse claudin-15[51] that 33-40 domain's secondary structure is not defined and 55-59 domain's secondary structure is coil. Claudin-15 hydrogen bond was 166±3 during 100ns. Both claudin-2 and -15's C-terminals showed high flexibility and were removed for the following molecular dynamics simulation as discussed before.

As a consequence, our results show homology modeling results claudin-1, -2, -15 are properly built and the structures are reasonable and stable enough for following molecular dynamics. Besides, we also found 194 amino acid domain for these three types of claudin is an ideal model for improving computational efficiency without significant impacts.

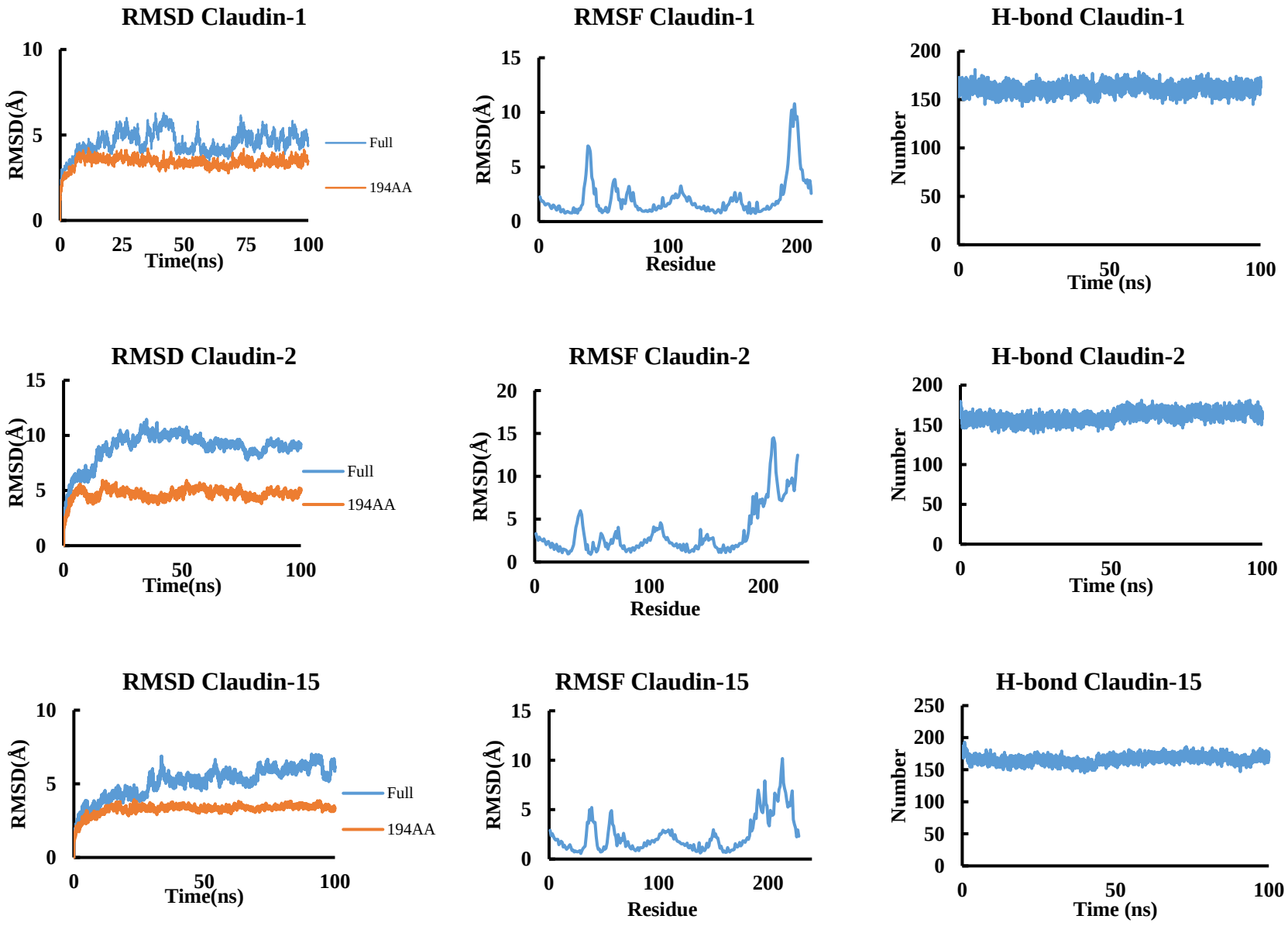


Figure 18 RMSD, RMSF and H-bond of claudin-1, -2, -15

3.5.2 Validation lipid membrane composition

In order to study homo cis-interaction process, a proper CG system environment is required to be built correctly. CG models of claudin were built from the validated atomistic structures and average structure during the last 30 ns of trajectory which RMSD showing no obvious fluctuation.

To characterize if the membrane would affect claudins' cis-interaction, claudin-1, -2, 15 were oriented by CHARMM server[67], and hydrophobic thicknesses were calculated. Gibbs free energy was a parameter to represent stability of protein configuration which means the energy needed from unfolded state to native folded state under optimal conditions[68]. These four claudins' Gibbs free energy were all below 0 around -50 kcal/mol that all this four structures weere all stable enough to be used as beginning structure. Tilt angle is the angle of helix long axis corresponding to the membrane orientation which is correspond to the minimum Gibbs free energy. As we see here, although claudin-15 showed higher average tilt angle than the other three, four tilt angle were overall kept in same level, so it will be the similar condition for them to behave in cis-interaction.

	Depth/Hydrophobic Thickness(Å)	ΔGtransfer (kcal/mol)	Tilt angle
Claudin-1	31.4 ± 1.8	-47.4	16.± 5.°
Claudin-2	32.8 ± 3.1	-55.3	14.± 4.°
Claudin-5	36.6 ± 3.6	-51.3	18.± 1.°
Claudin-15	31.0 ± 1.5	-51.7	27.± 8.°

Table 6 Hydrophobic Thickness and gibbs free energy of claudins

In real world, biological membrane are very dynamic and complex with various sorts of composition. And hydrophobic thickness, phase behavior, area per lipid, volume and changes in lipid structural order resulted from the composition variance will play important roles in

self-assemble process for transmembrane proteins[69]. If the lipid membrane thickness was largely below the protein hydrophobic thickness, protein will be tilted more or squeezed to fit the membrane thickness thus increasing the whole potential energy of system. If the lipid membrane thickness was largely above the protein hydrophobic thickness, protein will be stretched to lose their secondary structures or will be over disordered and cross-linked[70]. Also, the phase behavior also need to be taken into consideration that should be dynamic enough to be used in our computational simulation. It would make no sense for those simulation if the lipid membrane is too much sticky or rigid that proteins cannot diffuse at all. So, for demand of accuracy and efficiency, selecting a proper lipid composition to be used in molecular dynamics simulation would be a very critical step. DOPC, 1,2-di-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine, was selected to be the lipid membrane composition which thickness of bilayer is measured by stomic-force microscopy (AFM) to be around 31 Å at the 36°C with low temperature of melting(TM) at -17°C[71]. The hydrophobic thickness is properly fit the range of all four claudins' hydrophobic thickness. All four TM domains will not be tilted and squeezed too much that affects interaction among monomers. Low TM of DOPC make lipid membrane gel phase that protein can diffuse faster to improve computational efficiency without affecting cis-interaction.

3.5.3 Cluster analysis

Four types of claudin were conducted molecular dynamics simulation under same environment with random interfaces initially. Monomers began to interact with each other and aggregate to form macromolecular strands and Fig 19 shows 10 microseconds' results. The morphology of macromolecular strands looks similar to claudin based macromolecular strands observed in the freeze fracture microscopy images.

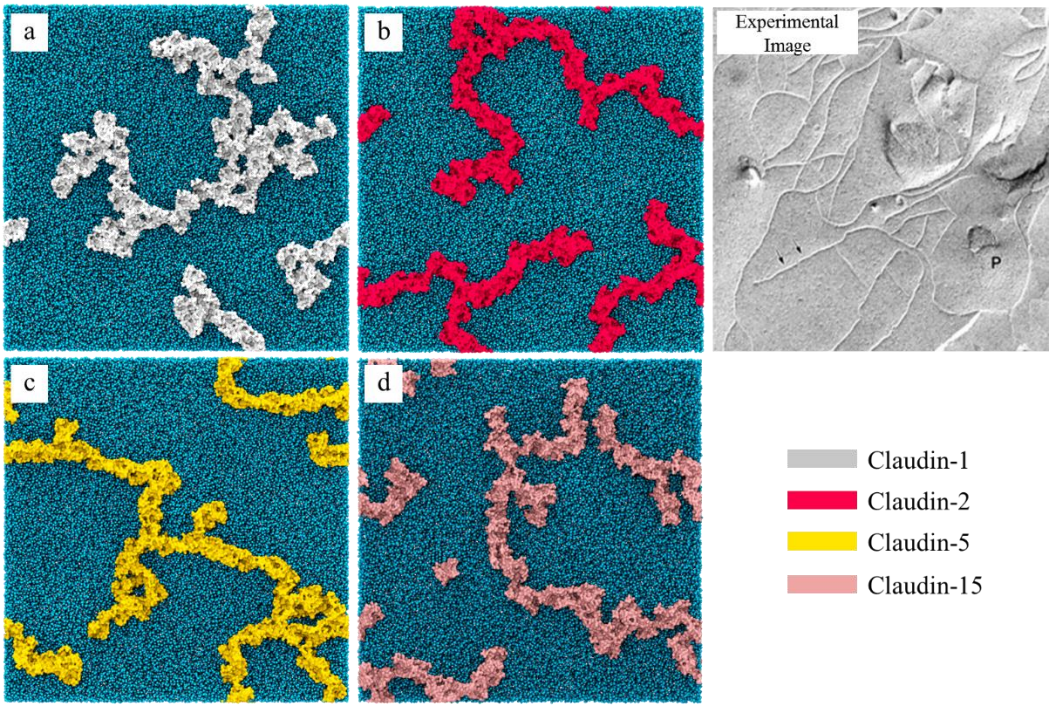


Figure 19 Snapshots of claudin macromolecular strand in MD simulation at 5 microsecond and freeze fracture image of Claudin-1 based macromolecular strand

Cluster analysis was conducted to the trajectory of each claudin to explore the cluster behavior pattern. Largest size of cluster and number of cluster was tracked in first two microseconds because no significant change observed afterwards. The claudin monomer's size on x and y axis is around 3 Å, and cut off to determine the cluster was set at 5 Å which means gap distance between two monomer below 2 Å would be define as interacted. The cutoff distance is kept consistent with FJ Irudayanathan's work about center of mass distance of different monomer in dimer oligomerization. Three repeat simulations were performed under same environment with random monomer interface orientations. Largest size of cluster and number of cluster plots was taken average based on three repeat simulations.

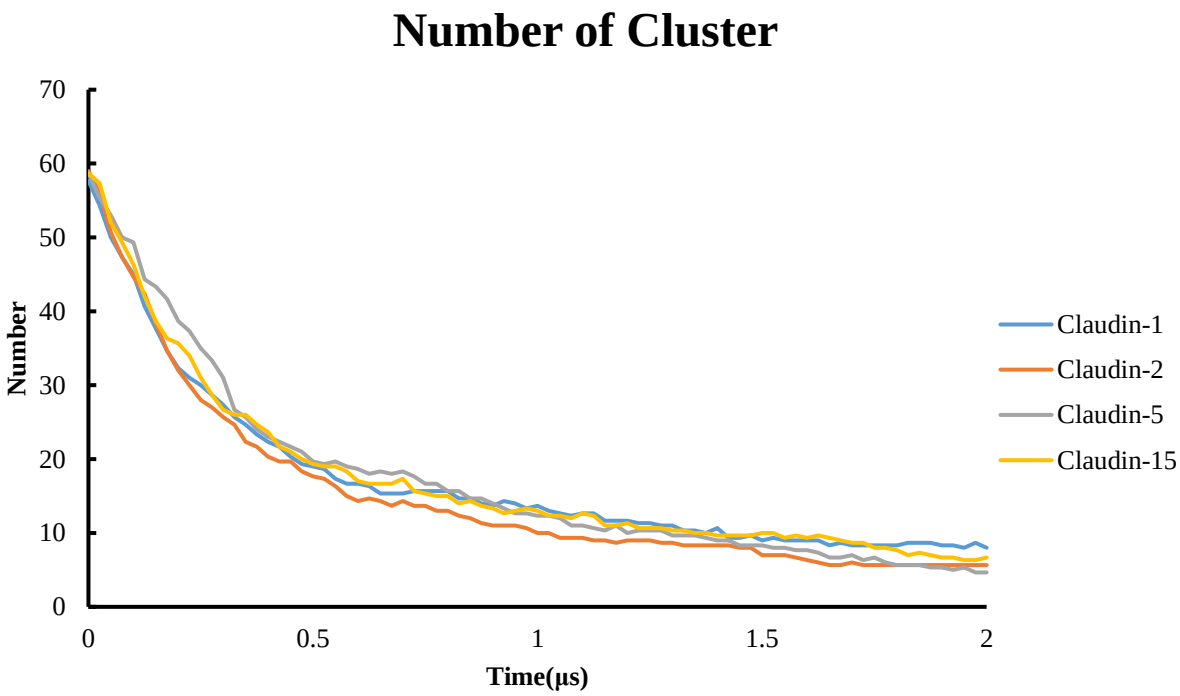


Figure 20 Number of cluster during 2 microseconds

The number of cluster shows change of number of cluster in system versus time. By tracking the number of cluster, difference in oligomerization speed could be shown. Claudin-1, -2, -5, -15 clusters numbers began from 58, 59, 59, and 59 respectively (Fig20). The numbers were not 64 because the interfaces were randomly oriented at the z axis of A87 backbone bead rather than center axis of monomer, thus some monomers' center of mass distance might be be below 5Å which was possible to be determined as interacted. But, since number of cluster of each cluster were very close at beginning and cutoffs were set as same, there would be little influence to see the difference of oligomerization speed during the whole trajectory. From the trajectory, all four claudins oligomerized in similar speed that interacted with each other and aggregate rapidly in the first 0.5 microsecond, then the speeds slowed down continuously. It is reasonable to understand the oligomerization process behavior that the larger cluster would diffuse slower because of the larger mass. Although at the end of 2 microseconds, claudin-1, -2 -5 and -15 formed 8.00, 5.67, 4.67, 6.33 clusters on average respectively, the difference is so tiny that may

resulted from randomness of system. Overall, four types of claudin showed very similar cluster behavior pattern on oligomerization speed.

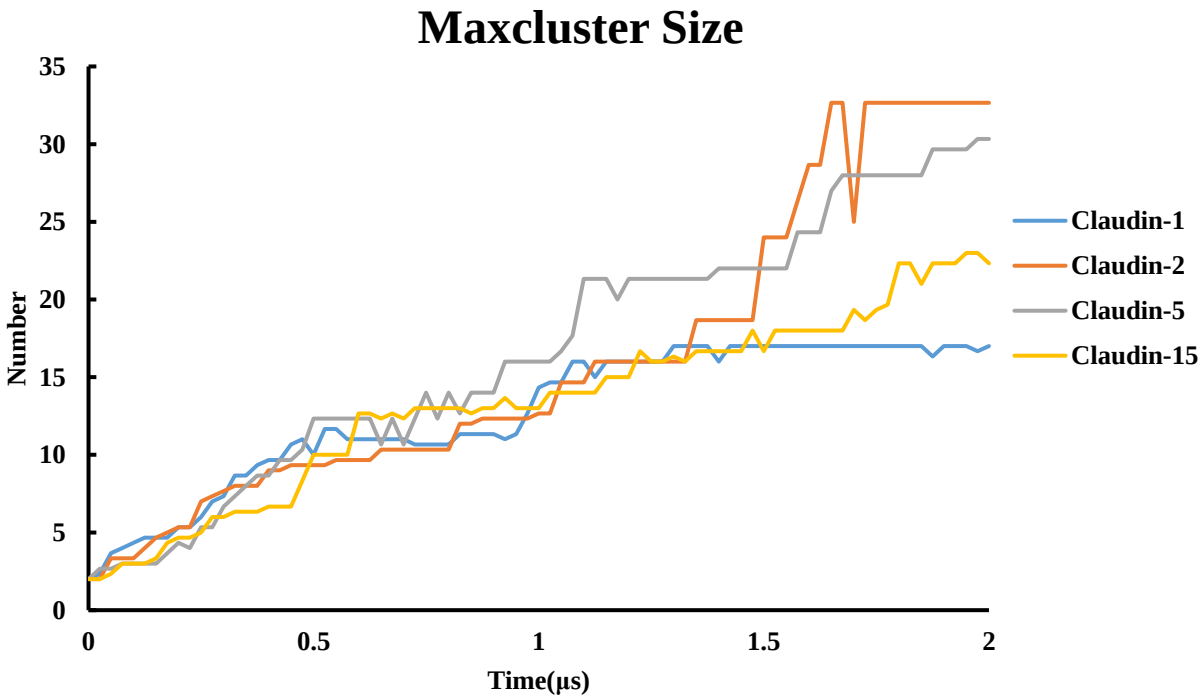


Figure 21 Maximum monomer number in cluster

The max cluster size shows maximum number of monomers aggregating into one cluster and it always shows the largest cluster. From the Fig20, Claudin-1, -2, -5, -15 began oligomerization at 2 monomer cluster which the reason has been mentioned before and formed 17.00, 32.67, 30.33, 22.33 monomer clusters at end of two microseconds. The results were taken average by three repeat simulations. For the reason that oligomerization speed began to slow down largely at 0.5 microsecond and numbers of cluster of each claudin remaining extremely similar after 1.5 microsecond, it is meaningless to track max cluster number after 1.5 microseconds because the huge change in max cluster size might because of aggregation by two middle size cluster in specific case. Therefore, although claudin-1 showed the smallest cluster by 17 monomer and claudin-2 formed 32.67 monomers cluster at the end of 2 microseconds.

In summary, from the results shown above, four types of claudin showed similar behavior pattern in both number of cluster and max cluster size.

3.5.4 Four classified dimer in simulation

In order to explore the details of strands, four unique dimer interfaces which have been observed consistently in four systems were classified and labeled from A-D. The A-D dimers samples shown in this session were picked from the 5 microseconds' simulation results. Dimer A and Dimer C was most observed in four systems. Dimer A is driven by both hydrophobic-hydrophilic interaction and electrostatic interactions. For the hydrophobic-hydrophobic interaction, dimer A is driven by hydrophobic amino acid on small helix in ECL1 and TM3*-ECL2* hydrophobic area. For electrostatic interaction, dimer A is formed by opposite charged amino acid on 4th β sheet on ECL1 domains and TM3*-ECL2*-TM4*. Snapshots of structure and contact map is shown as below. Although the interaction might be formed by various non-conserved amino acid for different claudin types, since these amino acid could show similar chemical interaction, same dimer interface configurations were shown in different claudin strands. Configuration analyses based on sequence alignment (Fig 22) and dimer configuration are shown below.

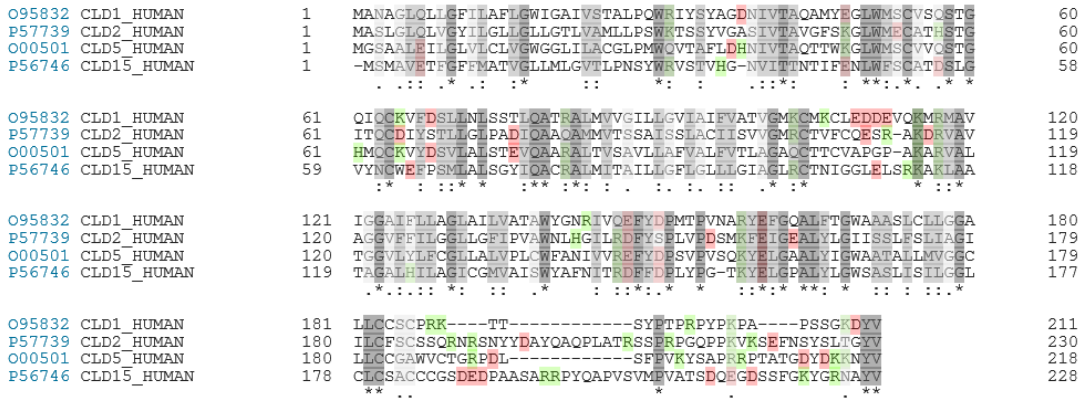


Figure 22 Sequence alignment of claudin-1, -2, -5, and -15. Grey highlight: Similarity, green highlight: positive, red highlight: negative

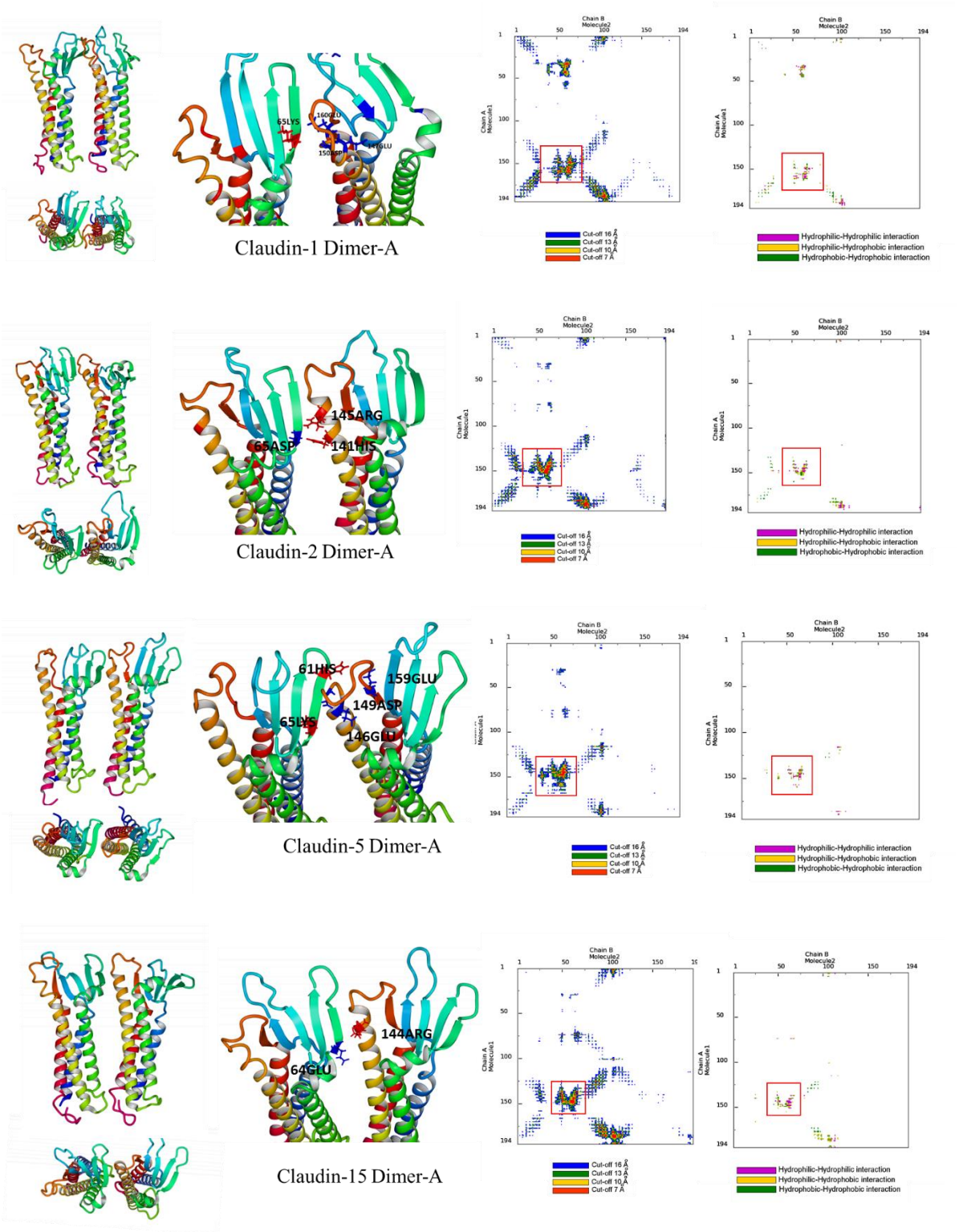
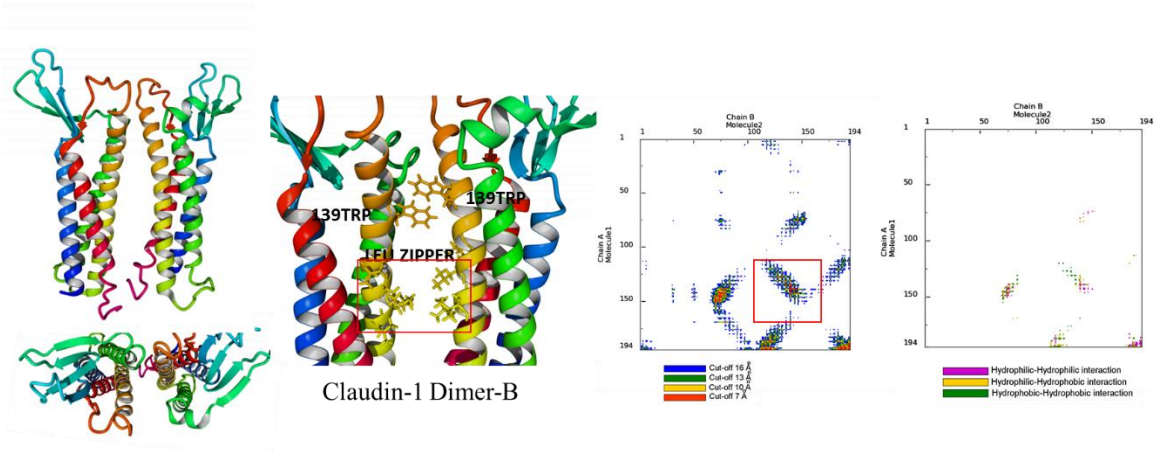


Figure 23 Detailed dimer structural analysis and contact area map of Dimer A

Dimer A hydrophobic-hydrophobic interactions is formed by conservative mutations amino acid which has similar chemical characteristics in those four claudins. Claudin-1, -2, -5 and -15 dimer A is formed between 70 LEU and 149PHE*-150TYR*, 70 LEU and 148PHE*-

149TYR*, 70 VAL and 148PHE*-149TYR*, and 68MET and 147PHE*-148TYR*, respectively. It is also showed that claudin-1 Dimer A is formed by positive charged 65LYS interacted with negative charged 147GLU-150ASP-160GLU cap. Claudin-2 Dimer A, oppose to claudin-1 that negative charged 65ASP interacted with positive 141HIS-145ARG162-GLU. Claudin-5 Dimer A is formed by two pairs of salt bridges, positive charged 65LYS interacted with negative charged 146GLU-149ASP and positive charged 61HIS interacted with negative charged 159GLU. Claudin-15 Dimer A is formed by negative charged 64GLU interacted with 144ARG. Although the interacted residues are not consistent, all four claudin ECL1 64-65 residue would anchored same ECL2-TM3 region with opposite charged. Besides, dimer A configuration in this study show agreement with the predicted dimer model in mouse claudin-15 crystal structure research that hydrophobic interaction between 68MET promoter and PHE146*-PHE147*-LEU157* promoter as well as some electrostatic interactions nearby would contribute to the dimer A configuration.



Claudin-1 Dimer-B

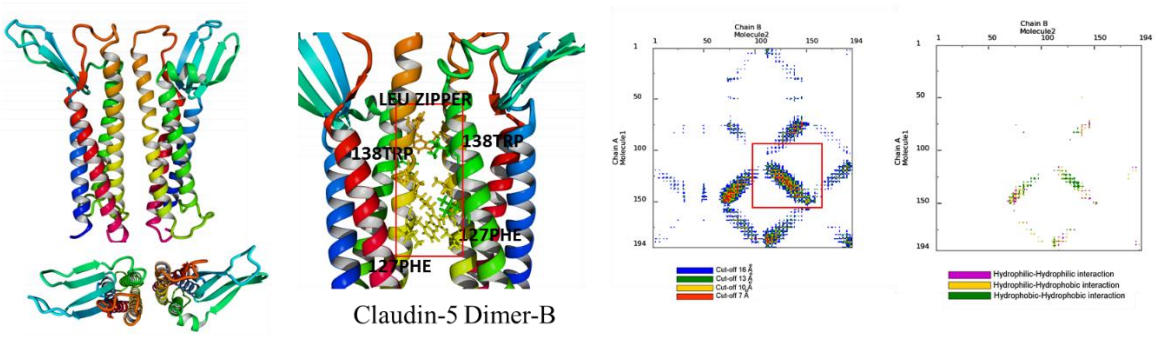
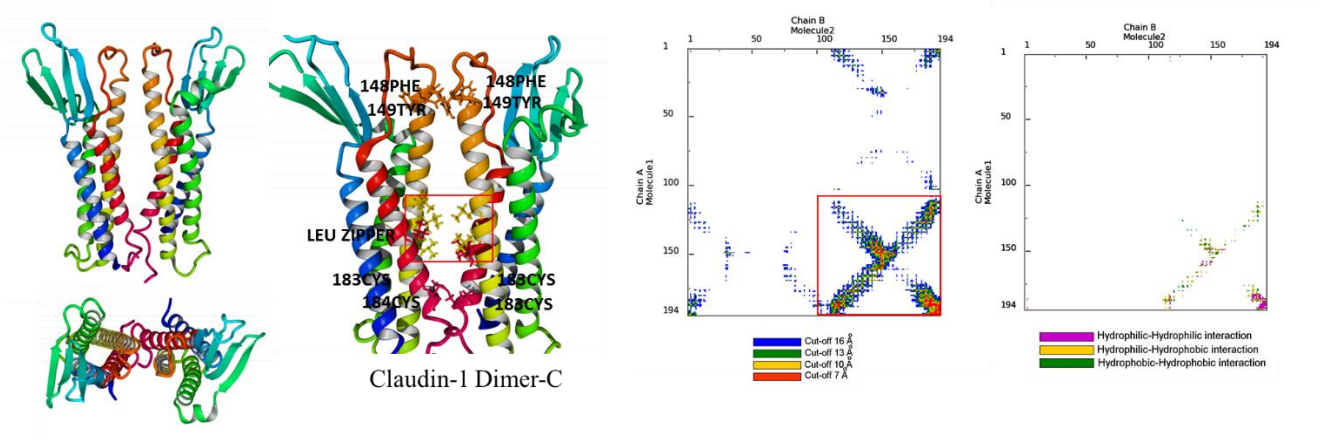
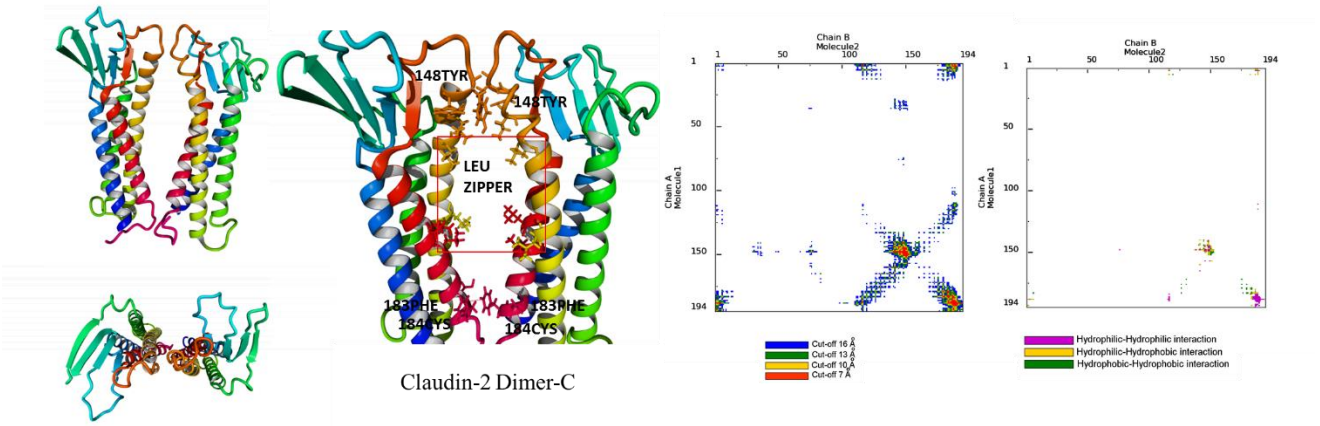


Figure 24 Detailed dimer structural analysis and contact area map of Dimer B

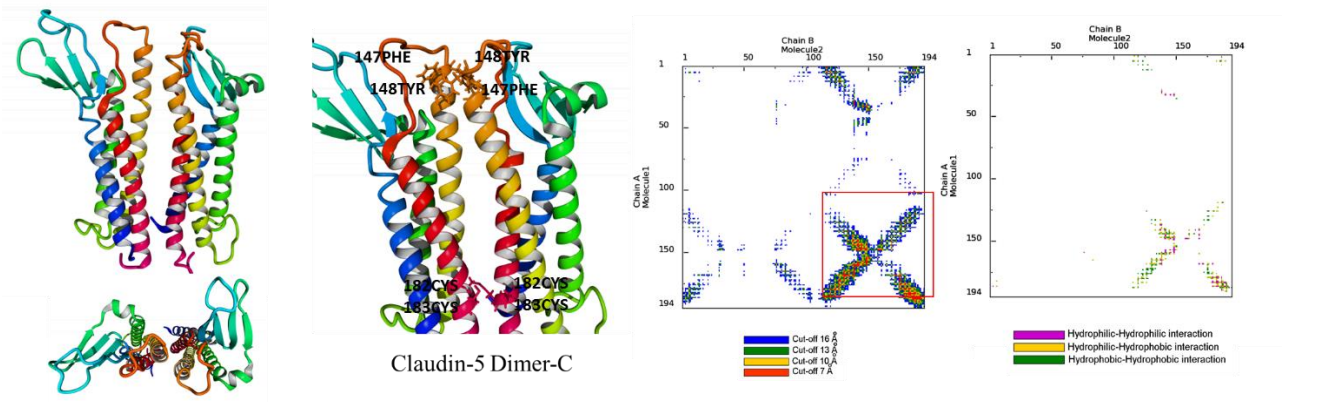
Dimer B is mainly mediated by TM3. However, it is rarely found in claudin-2 and -15 while it is seldom found in claudin-1 strand and prevalently found in claudin-5 strand. Dimer B in claudin-1 was formed by pairs of leucine zipper between TM3 and TM3* domains and 139TRP-139TRP Pi-Pi interactions. In Claudin-5 Dimer B, two monomers were dimerized by 138TRP-138TRP and 127PHE-127PHE Pi-Pi interactions and pairs of leucine zipper between TM3 and TM3*. By these interaction, a hydrophobic interface was built and make interaction stable. Leucine zippers were also found in Dimer C of all four claudins. Although leucine zipper is only reported be observed in FJ Irudayanathan's work that contribute to Dimer B and C configuration, it has been found to also contribute dimerization in erythropoietin, aquaporin, and discoidin domain family of receptor tyrosine kinases (DDR1 and DDR2). A high evolutionary conservation residues were found in these four TM domain protein means that it might play an important role in dimerization. These dimer B in two types of claudin were found to match predicted model in published paper that TM3 mediate the dimer B configuration.



Claudin-1 Dimer-C



Claudin-2 Dimer-C



Claudin-5 Dimer-C

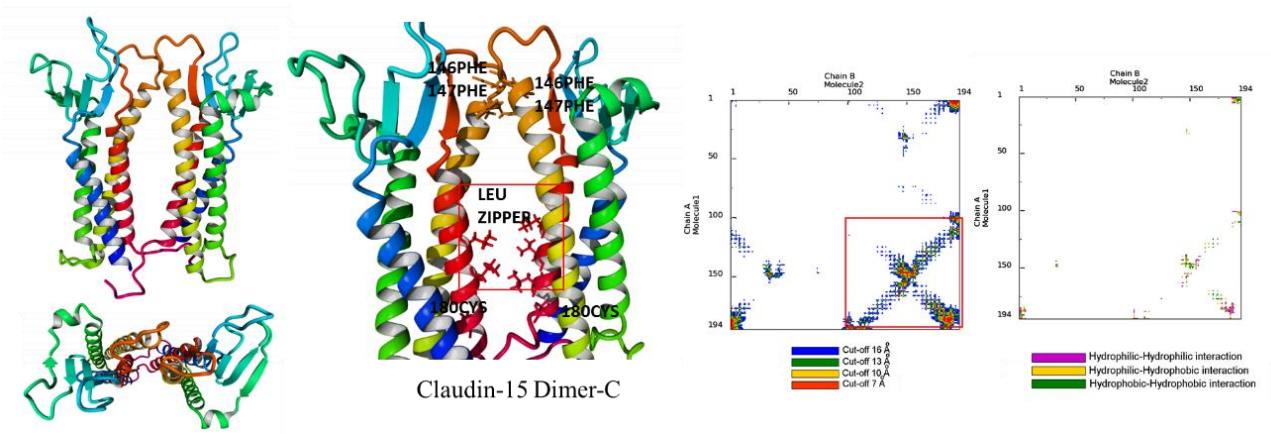


Figure 25 Detailed dimer structural analysis and contact area map of Dimer C

Dimer C is formed between TM3-ECL2-TM4 of two monomers. From four contact map, we are able to clearly see the closest area between two monomers is TM3-ECL2-TM4 part. This discovery keeps consistence with Yongfeng Gong et al's results about hetero cis-interaction model of claudin-16, -19 that TM3, TM4 and ECL2 are critical in cis- and trans- interaction[55b]. Despite experimental predicted models were produced by using different types of claudin, the residues proven to be critical in dimer C configuration were found to be conserved residues in most claudin or have similar chemical characteristics that might happen in most claudins. In that paper, by using alanine-insertion mutagenesis approach, they discovered several mutation points on TM3, and TM4 of claudin-16, and -19 not TM1 and TM2 would cause interaction loss. In our simulation, Dimer C configuration is similar with the model predicted in Yongfeng Gong's work. From the Fig 25, leucine zippers were found in all four types of claudin that might provide high binding affinity for dimer structures as discussed previously. Besides, two conserved residues CYS around 180 from TM4 could form cysteine-cysteine and hydrophilic-hydrophobic interactions between two monomer in dimer C configuration. Cysteine, as considered as a special amino acid, its thiol group and hydroxyl groups could greatly stabilize the dimerization by interacted with hydrophobic and hydrophilic region close to it such as LEU, ILE, PHE, and SER in four claudins[72]. Besides these interactions found in all claudin dimer C

configuration, Pi-Pi interaction on ELC2 ECL2 that 149TYR and 148PHE formed pi-pi stacking with another 148PHE and 149TYP to stabilize the Dimer C type. Some non-conserved residue might also contribute to the dimer C configuration. In claudin-2, Pi-Pi interaction occurred between 183PHE on TM4 of two different monomers. In claudin-5, 128CYS might provide additional binding affinity by hydrophobic-hydrophobic interaction. In claudin-15, 130 CYS might provide the same effect for dimer C configuration.

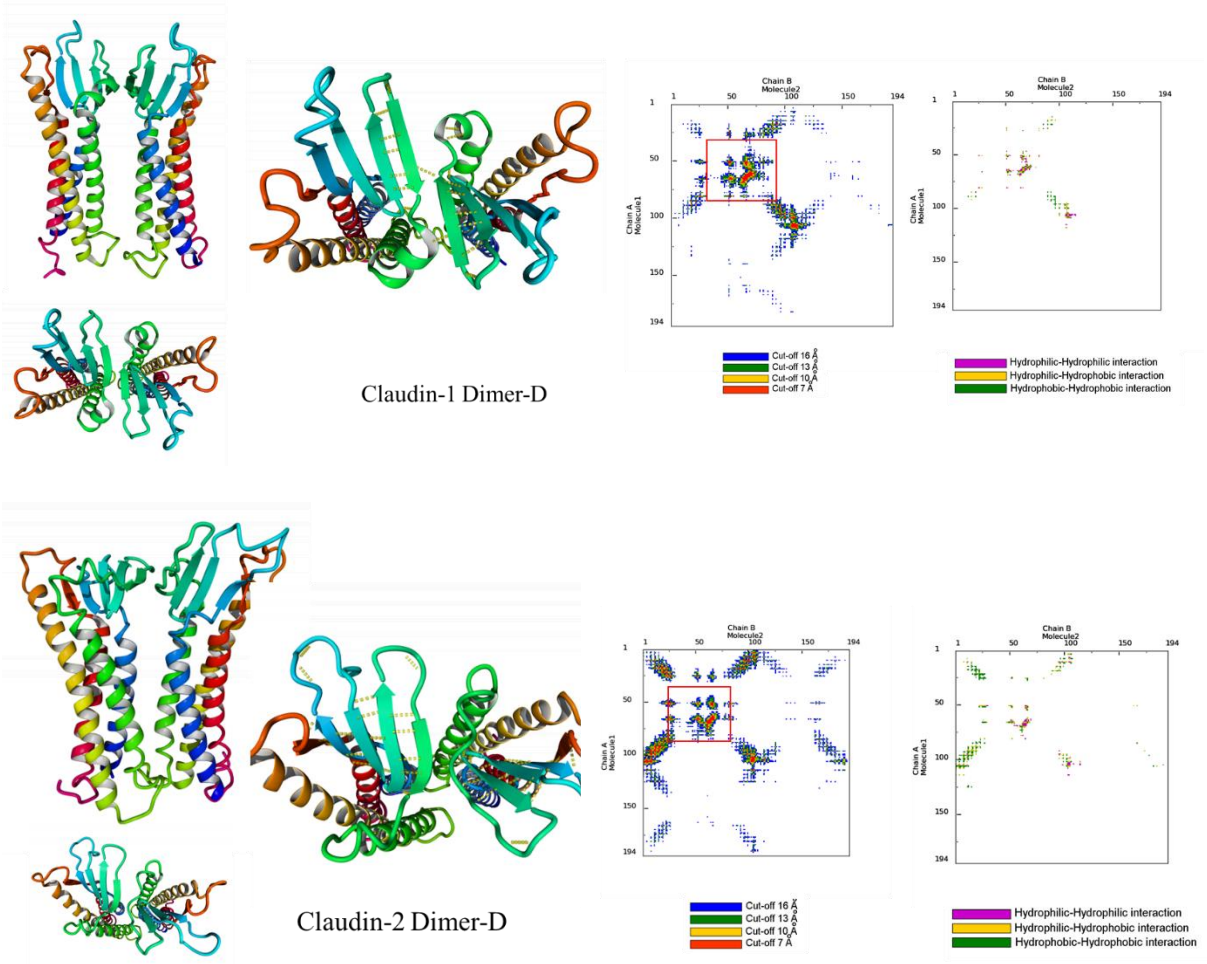


Figure 26 Detailed dimer structural analysis and contact area map of Dimer D

Dimer D configuration has been predicted by several research that claudin-1, -2, -3, and -15 could form dimer by anti-parallel β-sheet arrangement of ECL-1 residues and it is the critical dimer configuration for the channel function. Here our systems showed same configurations with those researches. The interaction was mainly driven by pairs of hydrogen bonds. Moreover,

the 65CYS and some hydrophobic amino acid like ILE, VAL, PHE around the 4th β-sheet which may leaded to hydrophobic-hydrophobic interaction with the same interface of another monomer. On the other hands, dimer D type occurred multiple times in claudin-1 and -2 while not found in claudin-5 and -15 at 5 microseconds. Although it happened at 2 microseconds frame snapshot (not shown here), it disappeared afterwards in 5 microseconds frame snapshots for both claudin-5 and -15. After tracking the trajectory, Dimer D anti-parallel β-sheet arrangement was shifted. It is possible that Dimer D type is not stable for claudin-5 and -15 that during the aggregation, thus interacted interfaces would change during MD simulations. FJ Irudayanathan's previous work[57] about claudin-5 cis-interaction also showed Dimer D type in claudin-5 would have lowest potential mean force which means the Dimer D interaction would be the least stable interaction among these four types of dimer.

So far, from the interactions discussed above, we showed the potential residues for all dimer type configuration. Our simulation systems also showed excellent agreement with the models has been discovered and were able to reproduce all models from current existing experiments. Even some models were built from mutation work in experiments, these models are still far away from elaborate the self-assembly process, and construction of TJs. Our MD simulations were able to, on the side, show atomistic level details about claudin cis-interaction and could be used as an assistance to guide the experiments.

3.5.5 Distribution of Dimer Type

Even the fact similar cluster pattern as well as similar basic dimer unit in the macromolecular strand, different detailed construction of strand might contribute in different types of claudin strand formation. So, we calculated the distribution of each dimer type by both hard distance criteria and soft angle criteria mentioned in *session 3.3.3 and 3.3.4* by using R programming

script. The determining R programming script was based on the distance matrix of TM distance and potential residues discussed in *3.5.4 session*. The distributions of dimer type are shown as below. All distributions were taken average from three repeat simulations at the end of 3 microseconds, and were shown in percentage rather than number because the sample sizes captured were different in different systems. The distribution of each dimer type is a very significant founding that current technology could not attach, and our work on counting first show the potential distribution.

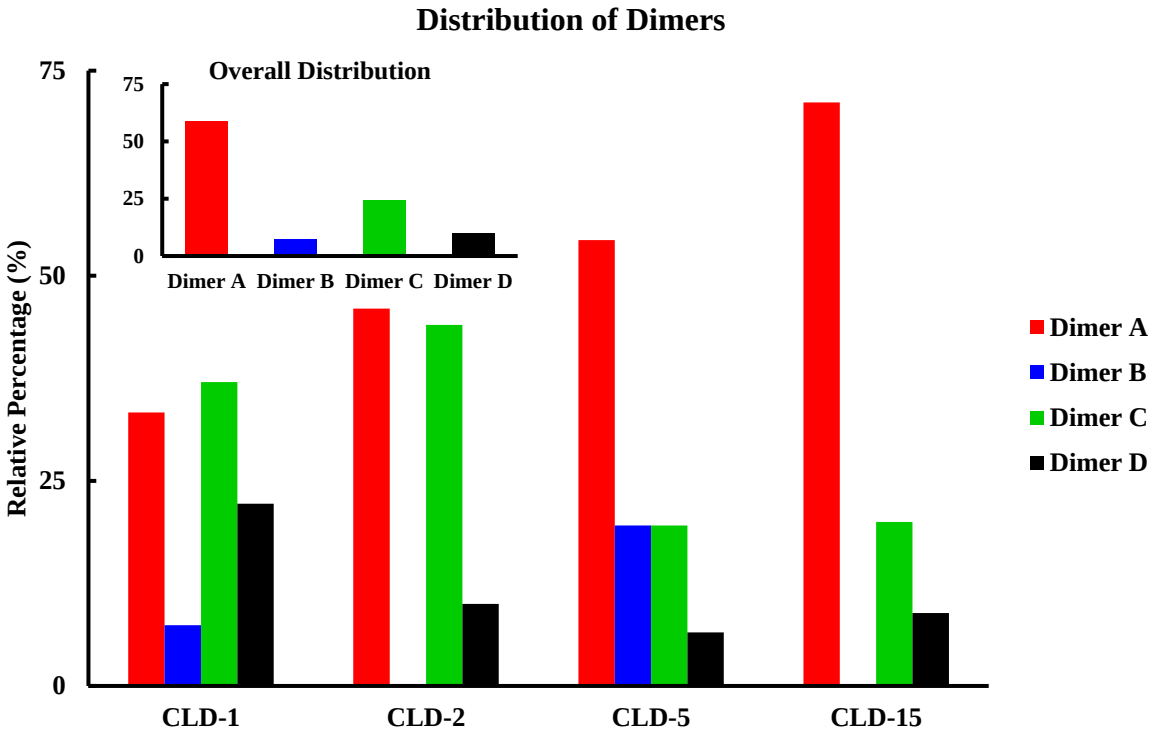


Figure 27 Distribution of four classified dimer type in claudin macromolecular strand

Fig 27 shows the distribution under hard distance criteria, the dimer counted here is exact dimer configuration without any variant configuration. Dimer A shows 58.6% in all claudin system and dimer A type shows the highest portion 46%, 54.35%, and 71.11% in claudin-2, -5, and -15, and second portion 33.33% in claudin-1. Dimer C, D is second and third number of dimer type counted in claudin-2, -5 and -15, while Dimer C is a little higher portion than dimer A in

claudin-1. This result is found to show agreement with one of the predicted dimer types in claudin-15 crystal structure. In that experimentation works, claudin strands were predicted as dimer width strand and were constructed by a pair of dimer A strand end to end. Dimer A was the unit to elongate the strand.

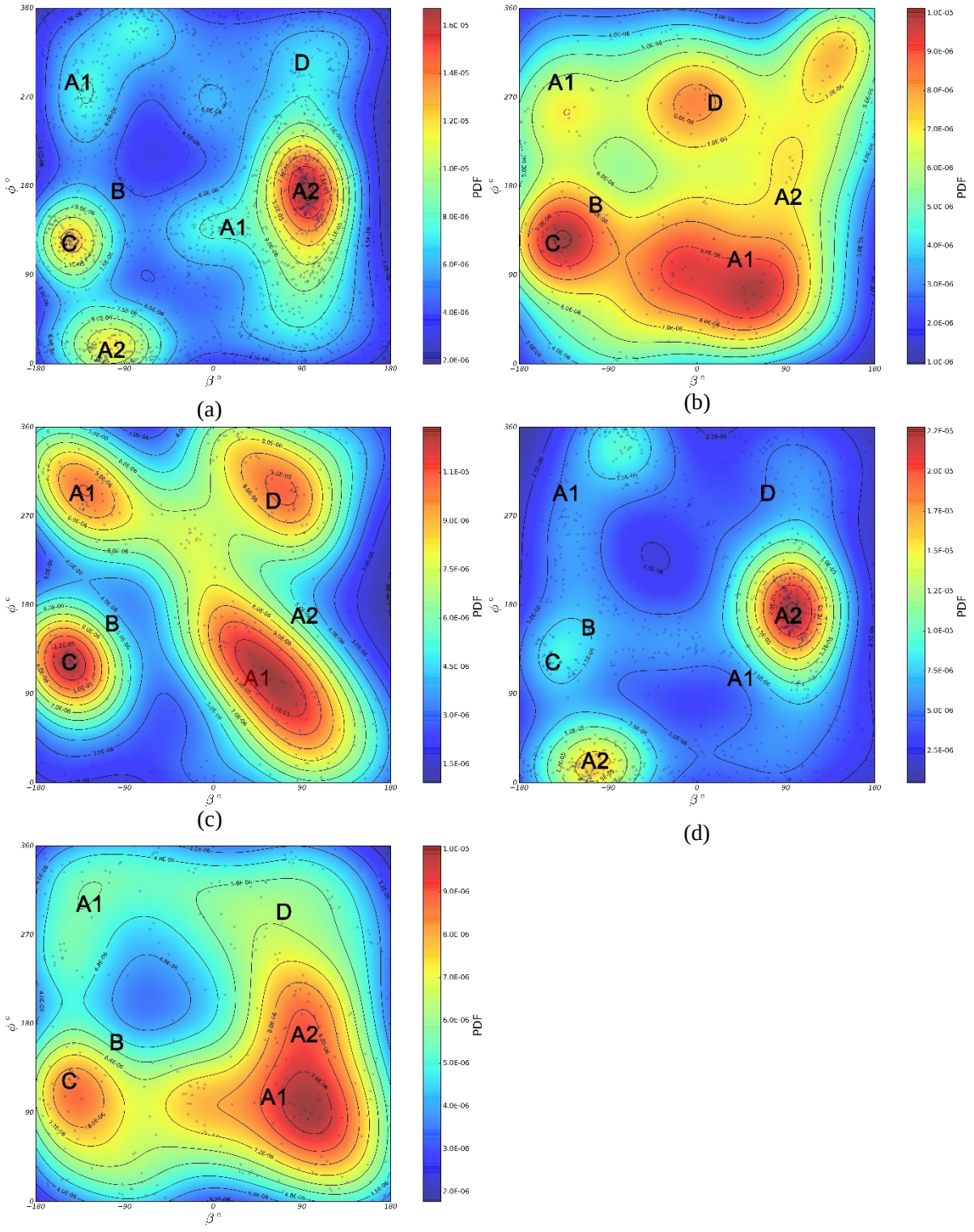
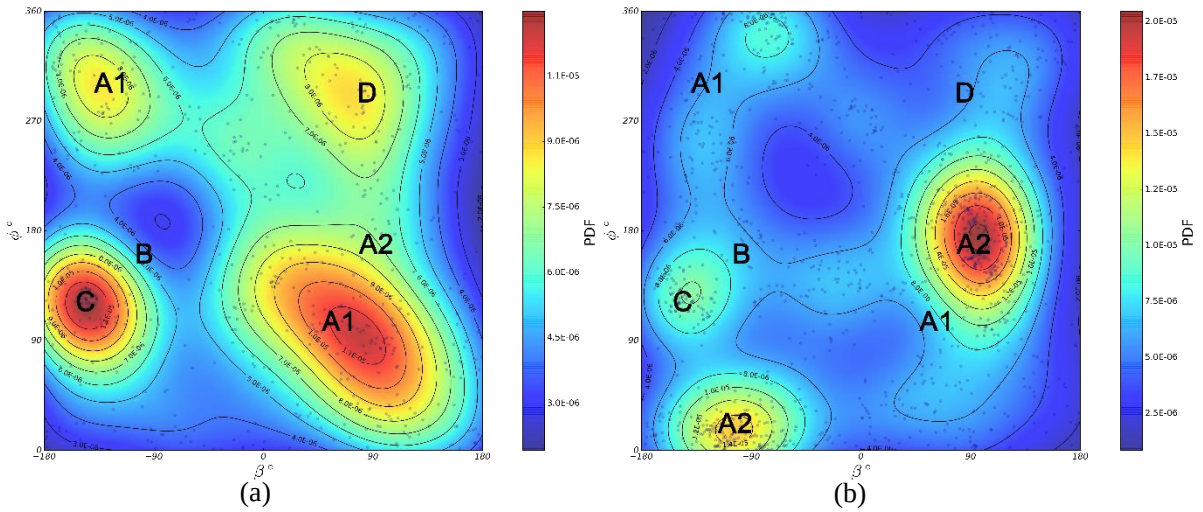


Figure 28 Probability density map of angle distribution in all four claudin systems. (a)Overall distribution (b) Claudin-1 distribution (c) Claudin-2 distribution (d) Claudin-5 distribution (e) Claudin-15 distribution

Since distributions above is based on hard distance criteria, it is hard to determine the variant of each dimer type and other unclassified dimer. So we used angle distribution to clarify all

dimer configuration Fig 28 shows the probability density map of angle distribution. On the contrary of hard distance criteria, angle distribution could not calculate the exact distribution of each dimer configuration but could show distinct dimer configuration and probability of variant which is similar to those four classified dimer configuration. So, probability density map of angle distribution were synergized with dimer distribution by hard distance criteria to explicit construction of claudin strands.

Dimer B, C and D is symmetric configuration so that they are shown in one spot area while dimer A which is asymmetric configuration is shown in two spot areas. It also needs to be mentioned that two variant was discussed in *session 3.5.4* that dimer A could be driven by hydrophobic-hydrophobic interaction as well as electrostatic interaction by various residues nearby β sheet in ECL1 domain. So, two spot areas were classified and used to represent dimer A configuration. As we shown in Fig 28, all high probability spot areas were classified in four classified dimer types and no obvious unclassified area occurred. Moreover, the probability keeps consistent with distribution in our hard distance criteria. All these probability density results validated our distribution calculated by hard distance criteria.



(a) (b)

Figure 29 Probability density map of angle distribution of pore forming and barrier forming claudin system. (a)Pore forming claudin (b) Barrier forming claudin

On the other hand, Dimer B configuration is only found in barrier forming claudin as claudin-1 and claudin-5 by hard distance criteria. Also, probability density map of angle distribution

show very low probability of dimer B configuration in barrier forming claudin. Moreover, dimer B configuration is unique in claudin-5 which was found to be the tie second highest dimer type in claudin-5 while was the least dimer type in claudin-1, and not shown in claudin-2 and -15. This result show it is possible dimer B is one unique dimer type for claudin-5 strand formation rather than the other claudin. On the other hand, according to the binding energy analysis in FJ's work, dimer B has the highest binding energy and dimer A shows the third highest binding energy among these four dimer types which means dimer B is energetic favorable configuration. But, it is found that dimer A was showed above two times more than dimer B. Furthermore, although four types of claudin share conserved residues to form leucine zipper and pi-pi interactions being able to form dimer B, pore forming claudin-2, and -15 shows few dimer B configuration. It is possible that the dimer configuration is driven by not only hydrophobic thickness but also the chemical characteristic of different lipid membrane composition. This result is expected to be studied further and proven by the future work.

In summary, it is significant that our *in silico* approach first time shows the distribution and numbers of distinct dimer type in strand construction and could be a guidance for future experimental works.

3.5.6 Mutation of Claudin-5

According to Jan Rossa's work[65] on claudin-5, I142T and F139S mutation was conducted in claudin-5, and linear claudin-5 strand formation was affected in experiments. In order to find out if dimer B could contribute in claudin-5 strand formation process, two amino acid mutation work was performed to claudin-5 as same as the experiments. Claudin-5 was manually mutated in sequence and simulated by same protocols as non-mutated system. All environment parameters used in MD simulation were kept same as non-mutated system except for amino acid sequence. For pursuing statistical significance, ten repeat works of both non-mutated

and mutated claudin-5 were done. Since both mutation amino acids are located on TM3 part, dimer A and dimer D is indirectly relevant to the mutation work so that change on dimer A and D is not discussed here. Both dimer B and C configurations are mediated by TM3, so that the change of dimer B and dimer C are emphasized here. All count number analyses are shown in Appendix.

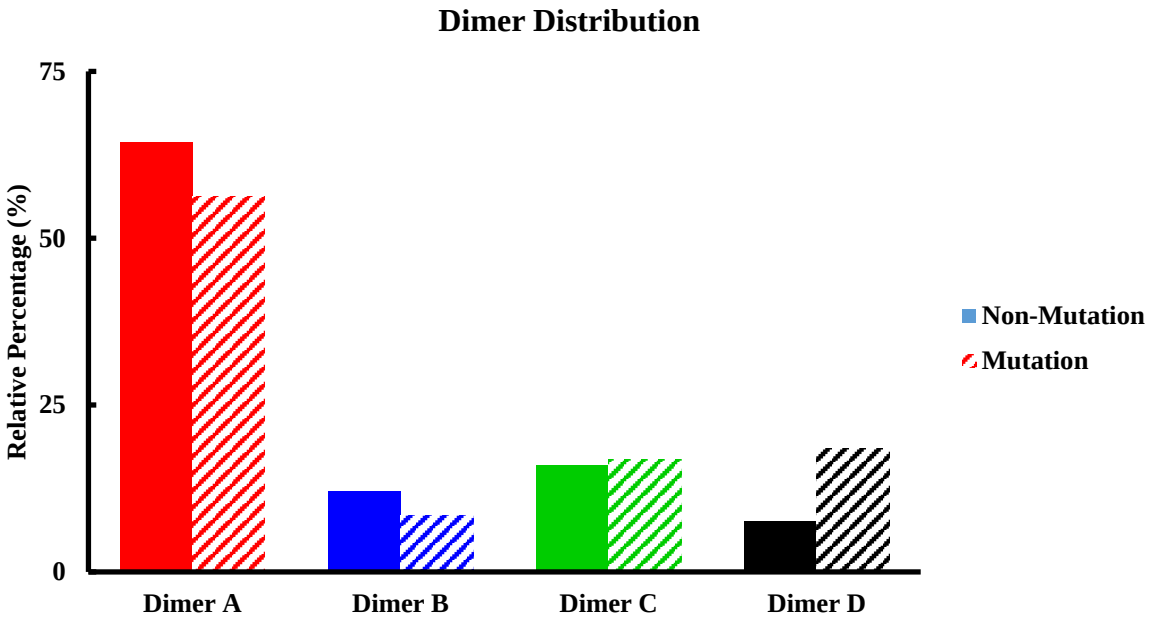


Figure 30 Distribution of dimer B type in non-mutated and mutated claudin-5 macromolecular strand

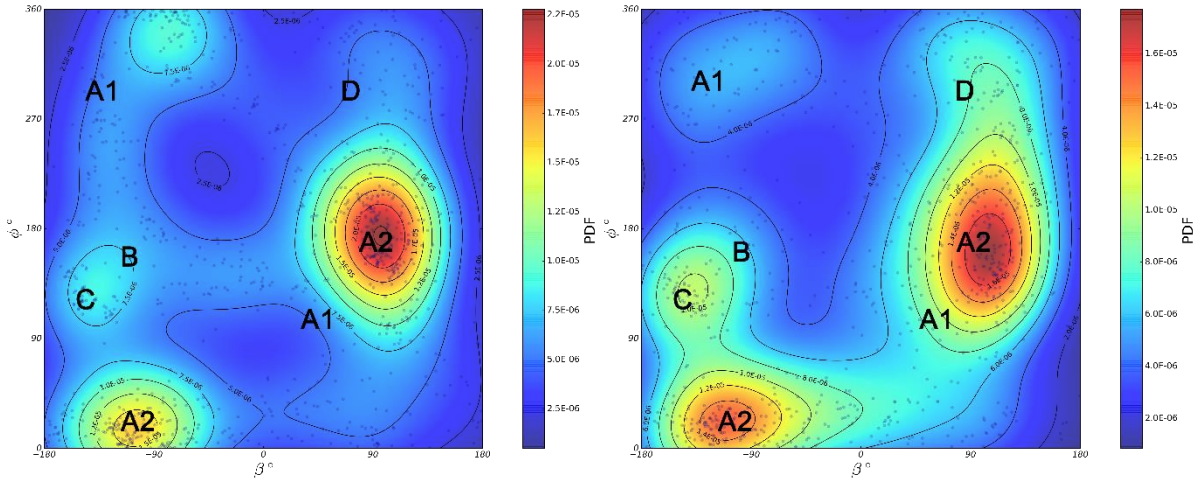


Figure 31 Probability density map of angle distribution of mutated and non-mutated claudin-5 system. (a)Pore forming claudin (b) Barrier forming claudin

From the Fig 28 shown, dimer A is the most predominate dimer type in claudin-5 strand in both non-mutated and mutated type. As these two amino acid mutation does not directly affect the dimer A configuration on both electrostatic and hydrophobic-hydrophobic interaction, decreasing on dimer A is probably resulted from the unstable binding affinity energy. The same condition was observed on dimer D that dimer D is not mediated by TM3 but increased by 11.91% after mutation. Dimer B and dimer C are both mediated by TM3 thus two amino acid mutation work would directly affects the configuration. After ten repeat 3 microseconds' work, dimer B decreased by 3.72% while dimer C increased by 0.9%. The number of total dimer B decreased from 16 to 10 while number of total dimer C increased from 20 to 21. Generally speaking, dimer C portion was kept in same level and dimer B portion decreased a little. Although change of dimer B portion is not obvious, it could be regarded as potential dimer type contributing strand formation because dimer B could be interface between two dimer A strands so that even small decrease would affect linear strand formation.

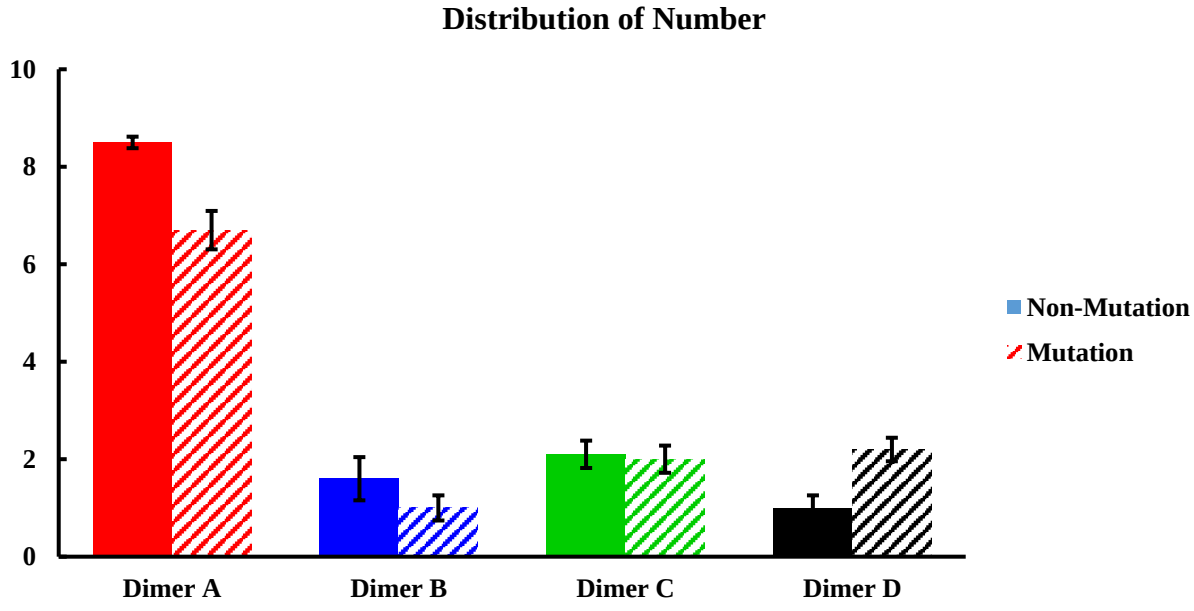


Figure 32 Number of dimer B type in non-mutated and mutated claudin-5 macromolecular strand

Through mutation work, we again showed the predominate portion of dimer A in

macromolecular strand, and dimer B rather than dimer C could be potentially critical dimer type in claudin-5 strand formation. Both of these detailed structural founding is significant finding because we first showed the change of complicated dimer distribution and they are extremely hard to be obtained by experiments.

3.5 Conclusion

Claudin is one of the most important protein in TJ that control the intercellular flows. Due to the complexity of experiments, the intact construction of TJ is still not revealed clearly. Although some mutation work and high-end techniques provide valuable information to help uncover the basic construction of TJs, the detailed construction is still elusive because of technique limitation. In this study, we used multiscale MD simulations approach to give new perspectives on the *cis*- interactions of claudin-5 TJs and show detailed architecture of claudin-5 TJs.

In order to study construction of TJ, 8*8 monomer grid with random interface was built on DOPC lipid membrane. After series of simulations, our 8*8 monomer grid CG models reproduce four dimer types same with experiments which is majority in strand formation. After discussed in details, these four dimer types showed excellent agreement with configurations in experiments. Two R-programming scripts were developed to calculate and determine the change of distributions of four classified dimer in our work, and here we first demonstrate the distribution of each dimer type in four claudin strand. Dimer A, C, D is found in all four claudin systems. Dimer A is the highest portion dimer type in most of claudin, while Dimer B is only highly expressed in barrier forming claudin -1 and -5. Also, by replicating the mutation work according to experiments, we showed dimer B is the potential dimer type in forming claudin-5 macromolecular strands. All these results could, on the side, provide a guidance to the experiments and an aid to further understand TJs and claudin monomers cis-interaction pattern.

4. Future work

Although this thesis gave an insight of molecules behavior in both nanocarrier and claudins, there are still lots of works needed to be done to further explain them in future. Molecular dynamics and application of CG models have been validated to be a great way in helping explore biological and chemical mechanism. More accurate force fields and more complicated models were built and used to help boost experimentation efficiency. However, there is still a long way from using simulation to perfectly reproduce the real experimentation, such as time-scale limitation and complexity limitation.

For telodendrimer nanocarrier area, we have built mature CG models and reproduced in vitro self-assemble process of three generations telodendrimers and doxirubincin and showed details of self-assemble behaviors. In future, more complicated environment is going to be used to mimic in vivo drug release process. pH factors can be in imported to simulate the environment of tumor mass or intestines and stomach by adding more charged ions to the environment thus these models can help develop drug delivery targeting ability or physical and chemical properties in oral administration. Also, fluid models can be used to simulate our drug to find out different rheological behaviors among different drug delivery pathways like blood stream. Moreover, novel dendrimer structures can be developed by revising or exchanging building blocks on our mature models.

For claudin area, four claudin models were built by homology modeling, four dimer types were classified and insight distribution models were given based on those classified models. However, limited research was done to find out the detailed constructions of TJs and structure-function relationship of claudins. In future, based on distribution of each claudins, more mutation works could be simulated to find out the relations between strand formation and each dimer type. Moreover, trans- interaction model or other TJ components like ZO and occludins could be applied to study the ion channel mechanisms and construction of TJs. Another

direction could be that add specific kind of drugs to our model and simulate the mass transport process. After all, there are still lots of things about claudin waiting to be solved nowadays such as mechanisms, constructions, biological properties, behavior in vivo and etc.

With higher requirements of atom level details in biological and chemical area, computational work has been increasingly needed and used to help provide insight and guidance to traditional experimentation. And with better understanding and application of MD simulation approach, great contribution would be provided to advance pharmaceutical, pathological and physiology areas.

5. Appendix

Script for determining numbers of each dimer type

Change pdb file to 7 colume format

```{r}
outcome_data <-
read.csv("C:/Users/xwang104/Dropbox/CLAUDIN_FAM/Xiaoyi/Dimer/cld15/csv/cld15_no
mut3_3usaa.csv",header = FALSE)

A<-
subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==17&outcome_data$
V2=="CA", select = c(V5, V6, V7)) #pick AA on TM1
B<-
subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==88&outcome_data$
V2=="CA", select = c(V5, V6, V7))#pick AA on TM2
C<-
subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==126&outcome_data$
V2=="CA", select = c(V5, V6, V7))#pick AA on TM3
D<-
subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==170&outcome_data$
V2=="CA", select = c(V5, V6, V7))#pick AA on TM4

trp<-
subset(outcome_data,outcome_data$V3=="TRP"&outcome_data$V4==137&outcome_data$
V2=="CA", select = c(V5, V6, V7))#138 TRP mid TM3
phe<-
subset(outcome_data,outcome_data$V3=="PHE"&outcome_data$V4==147&outcome_data$
V2=="CA", select = c(V5, V6, V7))#147 PHE top TM3
cys<-
subset(outcome_data,outcome_data$V3=="CYS"&outcome_data$V4==180&outcome_data$
V2=="SG", select = c(V5, V6, V7))#183 CYS down TM4
beta1<-
subset(outcome_data,outcome_data$V3=="ASN"&outcome_data$V4==61&outcome_data$
V2=="CA", select = c(V5, V6, V7))#mid 4beta sheet
beta2<-
subset(outcome_data,outcome_data$V3=="CYS"&outcome_data$V4==62&outcome_data$
V2=="CA", select = c(V5, V6, V7))#mid 4beta sheet
beta3<-
subset(outcome_data,outcome_data$V3=="TRP"&outcome_data$V4==63&outcome_data$
V2=="CA", select = c(V5, V6, V7))#mid 4beta sheet
betahead<-
subset(outcome_data,outcome_data$V3=="GLY"&outcome_data$V4==58&outcome_data$
V2=="CA", select = c(V5, V6, V7))#head ECL1 4beta sheet
betatail<-
subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==69&outcome_data$
V2=="CA", select = c(V5, V6, V7))#tail ECL1 4beta sheet
```

```

Calculate distance between All TM1 and TM1
```{r}
i <- 1:64; j <- 1:64;    #8*8 grid, can be other number
ni<- length(i); nj <- length(j)
y11 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni){
   for (p in 1:nj){
   y11 [p, k] <- dist(rbind(A[k,1:3],A[p,1:3]))
   }
}
```

Calculate distance between All TM1 and TM2
```{r}
y12 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni){
   for (p in 1:nj){
   y12 [p, k] <- dist(rbind(A[k,1:3],B[p,1:3]))
   }
}
```

Calculate distance between All TM1 and TM3
```{r}
y13 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
   for (p in 1:nj){
   y13 [p, k] <- dist(rbind(A[k,1:3],C[p,1:3]))
   }
}
```

Calculate distance between All TM1 and TM4
```{r}
y14 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
   for (p in 1:nj){
   y14 [p, k] <- dist(rbind(A[k,1:3],D[p,1:3]))
   }
}
```

Calculate distance between All TM2 and TM2
```{r}
y22 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
   for (p in 1:nj){
   y22 [p, k] <- dist(rbind(B[k,1:3],B[p,1:3]))

```
    }
}
```

Calculate distance between All TM2 and TM3
```{r}
y23 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
    for (p in 1:nj){
    y23 [p, k] <- dist(rbind(B[k,1:3],C[p,1:3]))
    }
}
```

Calculate distance between All TM2 and TM4
```{r}
y24 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
    for (p in 1:nj){
    y24 [p, k] <- dist(rbind(B[k,1:3],D[p,1:3]))
    }
}
```

Calculate distance between All TM3 and TM3
```{r}
y33 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
    for (p in 1:nj){
    y33 [p, k] <- dist(rbind(C[k,1:3],C[p,1:3]))
    }
}
```

Calculate distance between All TM3 and TM4
```{r}
y34 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
    for (p in 1:nj){
    y34 [p, k] <- dist(rbind(C[k,1:3],D[p,1:3]))
    }
}
```

Calculate distance between All TM4 and TM4
```{r}
y44 <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
    for (p in 1:nj){
    y44 [p, k] <- dist(rbind(D[k,1:3],D[p,1:3]))
```

```
   }
}
```

######for detailed amino acid
```{r}
#138Trp-Trp    interaction
ytrp <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ytrp [p, k] <- dist(rbind(trp[k,1:3],trp[p,1:3]))
   }
}

#146Phe-Phe distance
yphe <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   yphe [p, k] <- dist(rbind(phe[k,1:3],phe[p,1:3]))
   }
}

#183Cys-Cys distance
ycys <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ycys [p, k] <- dist(rbind(cys[k,1:3],cys[p,1:3]))
   }
}
```
####for Dimer D
````{r}
######for 63GLN-65LYS interaction
ybeta13 <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ybeta13 [p, k] <- dist(rbind(beta1[k,1:3],beta3[p,1:3]))
   }
}

######for 64CYS-64CYS interaction
```{r}
ybeta22 <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ybeta22 [p, k] <- dist(rbind(beta2[k,1:3],beta2[p,1:3]))
   }
}

######for 65LYS-63GLN interaction

```
ybeta31 <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ybeta31 [p, k] <- dist(rbind(beta3[k,1:3],beta1[p,1:3]))
   }
}

####for distance between two 4beta sheet head and tail
ybetaht <- matrix(NA, nj, ni)
for (k in 1:ni) {
   for (p in 1:nj){
   ybetaht [p, k] <- dist(rbind(betahead[k,1:3],betatail[p,1:3]))
   }
}

####stdev of (head,mid,tail)beta
ybetasd <- matrix(NA, nj, ni) # initiate a matrix with nj rows and ni columns
for (k in 1:ni) {
   for (p in 1:nj){
     w<- c(ybetaht[p,k],ybeta22[p,k],ybetaht[k,p])
   ybetasd [p, k] <- sd(w)
   }
}
```

```{r}
#Dimer A
#Criteria
belowcut11A<-subset(y11>20&y11>0)
belowcut14A<-subset(y14<24&y14>0)
belowcut23A<-subset(y23<24&y23>0)
belowcut33A<-subset(y33>20)

#Orientation
z12<-y11-y22
z13<-y11-y33
a<-subset(abs(z12)<10)
b<-subset(abs(z13)<10)
dimerA<-
which(belowcut14A==TRUE&belowcut23A==TRUE&belowcut33A==TRUE&belowcut11
A==TRUE&a==TRUE&b==TRUE, arr.ind = TRUE)
```

```{r}
#Dimer B
#Criteria
belowcut11B<-subset(y11>30)
belowcut33B<-subset(y33<20&y33>0)
belowcutpheB<-subset(yphe>14)
belowcuttrpB<-subset(ytrp<13&ytrp>0)
```

```
dimerB<-
which(belowcut33B==TRUE&belowcut11B==TRUE&belowcutpheB==TRUE&belowcuttrp
B==TRUE, arr.ind = TRUE)
```

```{r}
#Dimer C
#Criteria
belowcut34C<-subset(y34<24&y34>0)
belowcut11C<-subset(y11>20)
belowcutcysC<-subset(ycys<17.5&ycys>5)
belowcuttrpC<-subset(ytrp<21&ytrp>5)
belowcutpheC<-subset(yphe<14)

dimerC<-
(which(belowcutcysC==TRUE&belowcutpheC==TRUE&belowcut11C==TRUE&belowcut3
4C==TRUE, arr.ind = TRUE))
````

```{r}
#Dimer D
#Criteria
belowcutbeta1<-subset(ybeta13<14&ybeta13>0)
belowcutbeta2<-subset(ybeta22<14&ybeta22>0)
belowcutbeta3<-subset(ybeta31<14&ybeta31>0)
belowsd<-subset(ybetasd<10&ybetasd>0)
dimerD<-
which(belowcutbeta1==TRUE&belowcutbeta2==TRUE&belowcutbeta3==TRUE&belowsd
==TRUE,arr.ind=TRUE)
````

```{r}
#remove repeat
nA<-matrix(dimerA,nrow(dimerA))
nA1=t(apply(nA,1,sort))
nA2=unique(nA1)
nB<-matrix(dimerB,nrow(dimerB))
nB1=t(apply(nB,1,sort))
nB2=unique(nB1)
nC<-matrix(dimerC,nrow(dimerC))
nC1=t(apply(nC,1,sort))
nC2=unique(nC1)
nD<-matrix(dimerD,nrow(dimerD))
nD1=t(apply(nD,1,sort))
nD2=unique(nD1)

print(paste("A",nrow(nA2),"B",nrow(nB2),"C",nrow(nC2),"D",nrow(nD2))) ##final reuslts
````

Script to calculate external and internal angle within dimer

```
setwd("C:/Users/lenovo/Dropbox/CLAUDIN_FAM/Xiaoyi/RprogrammingCode/angle/angle/cld5/")
files<-dir()
for(m in files){
   outcome_data<-read.csv(m,header = FALSE)

   #outcome_data <-read.csv("C:/Users/xwang104/Dropbox/CLAUDIN_FAM/Xiaoyi/10repeat/csv/nomut/cld5_nomut6_3usaa.csv",header = FALSE)#from campus
   #outcome_data <-read.csv("C:/Users/lenovo/Dropbox/CLAUDIN_FAM/Xiaoyi/10repeat/csv/nomut/cld5_nomut1_3usaa.csv",header = FALSE)#home
   A<-subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==15&outcome_data$V2=="CA", select = c(V5, V6)) #pick AA on TM1
   B<-subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==90&outcome_data$V2=="CA", select = c(V5, V6))#pick AA on TM2
   C<-subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==126&outcome_data$V2=="CA", select = c(V5, V6))#pick AA on TM3
   D<-subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==173&outcome_data$V2=="CA", select = c(V5, V6))#pick AA on TM4
   beta2<-subset(outcome_data,outcome_data$V3=="CYS"&outcome_data$V4==64&outcome_data$V2=="CA", select = c(V5, V6, V7))#mid 4beta sheet
   betahead<-subset(outcome_data,outcome_data$V3=="GLY"&outcome_data$V4==60&outcome_data$V2=="CA", select = c(V5, V6, V7))#head ECL1 4beta sheet
   betatail<-subset(outcome_data,outcome_data$V3=="LEU"&outcome_data$V4==71&outcome_data$V2=="CA", select = c(V5, V6, V7))#tail ECL1 4beta sheet

   center<-data.frame(subset((A+B+C+D)/4,select = c(V5,V6)))
   plot(center,col="blue",pch=19,cex=2)
   text(center+10,labels=as.character(1:64))
   rownames(center)=1:64

   distmat<-data.matrix(dist(center))
   dimer<-subset(distmat<30&distmat>0)
   dimerPair<-which(dimer==TRUE,arr.ind=TRUE)

   i <- 1:64; j <- 1:64;    #8*8 grid, can be other number
   ni<- length(i); nj <- length(j)

   ybetaht <- matrix(NA, nj, ni)
   for (k in 1:ni) {
      for (p in 1:nj){
```

```
        if (k==p){
          next
        }
        ybetaht [p, k] <- dist(rbind(betahead[k,1:3],betatail[p,1:3]))
      }
    }
    belowcutbeta1<-subset(ybetaht<15&ybetaht>0)
    dimerD1<-which(belowcutbeta1==TRUE,arr.ind = TRUE)

    ybeta22 <- matrix(NA, nj, ni)
    for (k in 1:ni) {
      for (p in 1:nj){
        if (k==p){
          next
        }
        ybeta22 [p, k] <- dist(rbind(beta2[k,1:3],beta2[p,1:3]))
      }
    }
    belowcutbeta2<-subset(ybeta22<15&ybeta22>0)
    dimerD2<-which(belowcutbeta2==TRUE,arr.ind=TRUE)


    n<-matrix(rbind(dimerPair,dimerD2,dimerD1),nrow(rbind(dimerPair,dimerD1,dimerD2)))
    n1=t(apply(n,1,sort))
    n2=unique(n1)
    n3=n2[ , c(2,1)]
    n4=rbind(n2,n3)


    e<-1:nrow(n4);f<-1:2
    ne<-length(e);nf<-length(f)

    anglematrix<-matrix(NA,ne,5)

    for (h in 1:ne){
      a=n4[h,1]
      b=n4[h,2]

      c1=center[a,]
      c2=center[b,]
      t1=A[a,c(1:2)]
      t2=A[b,c(1:2)]

      ##angle for beta in mol A range (-180,180)
      v1a=t1-c1 #vector c1t1 not use
      v2a=c2-c1 #vector c1c2
      v3a=c(v1a[1,2],-v1a[1,1]) #vector verticle to c1t1 in clockwise rotation
      a2=atan2(((v3a[1]*v2a[1,2])-
(v3a[2]*v2a[1,1])),((v3a[1]*v2a[1,1])+(v3a[2]*v2a[1,2])))*180/pi  #beta  clockwise    Vector
OX to C1C2 use
```

```
      #angle for fi in mol B (0,360)
      v1b=t2-c2 #vector c2t2
      v2b=c1-c2 #vector c2c1
      b1=atan2(((v1b[1,1]*v2b[1,2])-
(v1b[1,2]*v2b[1,1])),((v1b[1,1]*v2b[1,1])+(v1b[1,2]*v2b[1,2])))*180/pi #fi

      if (b1<0) {
         b1=360+b1
      }

      anglematrix[h,1]=a
      anglematrix[h,2]=b
      anglematrix[h,3]<-distmat[a,b]
      anglematrix[h,4]<-a2
      anglematrix[h,5]<-b1
   }
   plot(anglematrix[,4:5])

   test<-kmeans(anglematrix[,4:5],centers = 5)
   plot(anglematrix[,4:5], col = test$cluster,pch=19,cex=1)


   outputfilepath<-"C:/Users/lenovo/Dropbox/CLAUDIN_FAM/Xiaoyi/angledistribution3/"
   #print(anglematrix[,4:5])
   write.csv(anglematrix,paste(outputfilepath,m,".csv",sep=""))

}
```

Dimer type count and percentage calculation from 3 repeat self-assembly simulations

claudin-1	Dimer A	Dimer B	Dimer C	Dimer D
1st	3	1	6	3
2nd	3	0	2	0
3rd	3	1	2	3

claudin-2	Dimer A	Dimer B	Dimer C	Dimer D
1st	7	0	10	2
2nd	9	0	6	1
3rd	7	0	6	2

claudin-5	Dimer A	Dimer B	Dimer C	Dimer D
1st	9	6	2	0

2nd	8	2	4	2
3rd	8	1	3	1

claudin-15	Dimer A	Dimer B	Dimer C	Dimer D
1st	12	0	3	1
2nd	8	0	3	1
3rd	12	0	3	2

Percentage	Dimer A	Dimer B	Dimer C	Dimer D
claudin-1	33.33%	7.41%	37.04%	22.22%
claudin-2	46.00%	0.00%	44.00%	10.00%
claudin-5	54.35%	19.57%	19.57%	6.52%
claudin-15	71.11%	0.00%	20.00%	8.89%

Dimer type count and percentage calculation from 10 repeat non-mutated and mutated self-assembly simulations

Non-mutated Claudin-5

Time	Dimer A	Dimer B	Dimer C	Dimer D
1	9	6	2	0
2	8	2	4	2
3	8	1	3	1
4	10	1	1	1
5	10	1	0	0
6	9	1	2	1
7	8	0	2	2
8	7	3	4	2
9	7	1	2	1
10	9	0	1	0

Muted Claudin-5

Time	Dimer A	Dimer B	Dimer C	Dimer D
1	4	0	1	3
2	4	2	2	1
3	10	1	2	0
4	4	0	3	2
5	8	1	4	2
6	9	2	1	4
7	3	1	1	3
8	12	1	4	2
9	4	2	1	2
10	9	0	1	3

Average Number	Dimer A	Dimer B	Dimer C	Dimer D
Non-mutated	8.5	1.6	2.1	1
Mutated	6.7	1	2	2.2

Ratio	Dimer A	Dimer B	Dimer C	Dimer D
Non-mutated	64.39%	12.12%	15.91%	7.58%
Mutated	56.30%	8.40%	16.81%	18.49%

6. Reference

1. Hansson, T.; Oostenbrink, C.; van Gunsteren, W., Molecular dynamics simulations. *Current opinion in structural biology* **2002,** *12* (2), 190-196.
2. Haile, J., *Molecular dynamics simulation*. Wiley, New York: 1992; Vol. 18.
3. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC biology* **2011,** *9* (1), 71.
4. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y., Atomic-level characterization of the structural dynamics of proteins. *Science* **2010,** *330* (6002), 341-346.
5. Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology* **2002,** *9* (9), 646-652.
6. Barenholz, Y. C., Doxil®—the first FDA-approved nano-drug: lessons learned. *Journal of Controlled Release* **2012,** *160* (2), 117-134.
7. Günzel, D.; Alan, S., Claudins and the modulation of tight junction permeability. *Physiological reviews* **2013,** *93* (2), 525-569.
8. (a) Nishiyama, N.; Kataoka, K., Current state, achievements, and future prospects of polymeric micelles as nanocarriers for drug and gene delivery. *Pharmacology & therapeutics* **2006,** *112* (3), 630-648; (b) Peer, D.; Karp, J. M.; Hong, S.; Farokhzad, O. C.; Margalit, R.; Langer, R., Nanocarriers as an emerging platform for cancer therapy. *Nature nanotechnology* **2007,** *2* (12), 751-760.
9. Nicolas, J.; Mura, S.; Brambilla, D.; Mackiewicz, N.; Couvreur, P., Design, functionalization strategies and biomedical applications of targeted biodegradable/biocompatible polymer-based nanocarriers for drug delivery. *Chemical Society Reviews* **2013,** *42* (3), 1147-1235.
10. Kamaly, N.; Xiao, Z.; Valencia, P. M.; Radovic-Moreno, A. F.; Farokhzad, O. C., Targeted polymeric therapeutic nanoparticles: design, development and clinical translation. *Chemical Society Reviews* **2012,** *41* (7), 2971-3010.
11. Milatz, S.; Piontek, J.; Schulzke, J.-D.; Blasig, I. E.; Fromm, M.; Günzel, D., Probing the cis-arrangement of prototype tight junction proteins claudin-1 and claudin-3. *Biochemical Journal* **2015,** *468* (3), 449-458.
12. Farquhar, M. J.; Hu, K.; Harris, H. J.; Davis, C.; Brimacombe, C. L.; Fletcher, S. J.; Baumert, T. F.; Rappoport, J. Z.; Balfe, P.; McKeating, J. A., Hepatitis C virus induces CD81 and claudin-1 endocytosis. *Journal of virology* **2012,** *86* (8), 4305-4316.
13. Nelson, P. T.; Alafuzoff, I.; Bigio, E. H.; Bouras, C.; Braak, H.; Cairns, N. J.; Castellani, R. J.; Crain, B. J.; Davies, P.; Del Tredici, K., Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *Journal of Neuropathology & Experimental Neurology* **2012,** *71* (5), 362-381.
14. Cabezas, R.; Ávila, M.; Gonzalez, J.; El-Bachá, R. S.; Báez, E.; García-Segura, L. M.; Coronel, J. C. J.; Capani, F.; Cardona-Gomez, G. P.; Barreto, G.

E., Astrocytic modulation of blood brain barrier: perspectives on Parkinson's disease. *Frontiers in cellular neuroscience* **2015,** *8*.
15. Krieger, E.; Nabuurs, S. B.; Vriend, G., Homology modeling. *Methods of biochemical analysis* **2003,** *44*, 509-524.
16. (a) Hooft, R. W.; Sander, C.; Vriend, G., Objectively judging the quality of a protein structure from a Ramachandran plot. *Computer applications in the biosciences: CABIOS* **1997,** *13* (4), 425-430; (b) Zhang, Y.; Skolnick, J., Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics* **2004,** *57* (4), 702-710.
17. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995,** *91* (1), 1-41.
18. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The REd. Tools: Advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics* **2010,** *12* (28), 7821-7839.
19. (a) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell Jr, A. D., Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *Journal of chemical theory and computation* **2011,** *8* (1), 348-362; (b) van Gunsteren, W. F.; Daura, X.; Mark, A. E., GROMOS force field. *Encyclopedia of computational chemistry* **1998**.
20. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.
21. Müller‐Plathe, F., Coarse‐graining in polymer simulation: from the atomistic to the mesoscopic scale and back. *ChemPhysChem* **2002,** *3* (9), 754-769.
22. Voter, A. F.; Montalenti, F.; Germann, T. C., Extending the time scale in atomistic simulation of materials. *Annual Review of Materials Research* **2002,** *32* (1), 321-346.
23. (a) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L., Coarse grain models and the computer simulation of soft materials. *Journal of Physics: Condensed Matter* **2004,** *16* (15), R481; (b) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H., The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **2007,** *111* (27), 7812-7824.
24. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J., The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* **2008,** *4* (5), 819-834.

25. Loverde, S. M.; Klein, M. L.; Discher, D. E., Nanoparticle Shape Improves Delivery: Rational Coarse Grain Molecular Dynamics (rCG - MD) of Taxol in Worm - Like PEG - PCL Micelles. *Advanced materials* **2012,** *24* (28), 3823-3830.
26. Zheng, M.; Pavan, G. M.; Neeb, M.; Schaper, A. K.; Danani, A.; Klebe, G.; Merkel, O. M.; Kissel, T., Targeting the blind spot of polycationic nanocarrier-based siRNA delivery. *ACS nano* **2012,** *6* (11), 9447-9454.
27. Singer, B.; Grunberger, D., *Molecular biology of mutagens and carcinogens*. Springer Science & Business Media: 2012.
28. Vipperla, B.; Febin Prabhu Dass, J.; Jayanthi, S., Insilico modeling and molecular dynamic simulation of claudin-1 point mutations in HCV infection. *Journal of Biomolecular Structure and Dynamics* **2014,** *32* (9), 1443-1455.
29. Bhavaniprasad, V.; Dass, J. F. P.; Jayanthi, S., Activation mechanism of claudin-4 by ephrin type-A receptor 2: a molecular dynamics approach. *Molecular BioSystems* **2013,** *9* (11), 2627-2634.
30. (a) Group, U. C. S. W., United States cancer statistics: 1999–2010 incidence and mortality web-based report. *Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute* **2013**; (b) Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2016. *CA: A cancer journal for clinicians* **2015**.
31. (a) Kroemer, G.; Galluzzi, L.; Kepp, O.; Zitvogel, L., Immunogenic cell death in cancer therapy. *Annual review of immunology* **2013,** *31*, 51-72; (b) Moore Jr, J. E.; Maitland, D. J., *Biomedical technology and devices*. CRC press: 2013; (c) Murtaza, M.; Dawson, S.-J.; Tsui, D. W.; Gale, D.; Forshew, T.; Piskorz, A. M.; Parkinson, C.; Chin, S.-F.; Kingsbury, Z.; Wong, A. S., Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **2013,** *497* (7447), 108-112; (d) Hacker, M.; Douple, E. B.; Krakoff, I. H., *Platinum Coordination Complexes in Cancer Chemotherapy: Proceedings of the Fourth International Symposium on Platinum Coordination Complexes in Cancer Chemotherapy Convened in Burlington, Vermont by the Vermont Regional Cancer Center and the Norris Cotton Cancer Center, June 22–24, 1983*. Springer Science & Business Media: 2012; Vol. 17.
32. Eisen, M., *Mathematical models in cell biology and cancer chemotherapy*. Springer Science & Business Media: 2013; Vol. 30.
33. Deng, C.; Jiang, Y.; Cheng, R.; Meng, F.; Zhong, Z., Biodegradable polymeric micelles for targeted and controlled anticancer drug delivery: promises, progress and prospects. *Nano Today* **2012,** *7* (5), 467-480.
34. Loi, S.; Sirtaine, N.; Piette, F.; Salgado, R.; Viale, G.; Van Eenoo, F.; Rouas, G.; Francis, P.; Crown, J. P.; Hitre, E., Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *Journal of clinical oncology* **2013**, JCO. 2011.41. 0902.

35. (a) Tacar, O.; Sriamornsak, P.; Dass, C. R., Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *Journal of Pharmacy and Pharmacology* **2013,** *65* (2), 157-170; (b) Von Hoff, D. D.; Layard, M. W.; Basa, P.; Davis, H. L.; Von Hoff, A. L.; Rozencweig, M.; Muggia, F. M., Risk factors for doxorubicin-lnduced congestive heart failure. *Annals of internal medicine* **1979,** *91* (5), 710-717.
36. Sparreboom, A.; Van Asperen, J.; Mayer, U.; Schinkel, A. H.; Smit, J. W.; Meijer, D. K.; Borst, P.; Nooijen, W. J.; Beijnen, J. H.; Van Tellingen, O., Limited oral bioavailability and active epithelial excretion of paclitaxel (Taxol) caused by P-glycoprotein in the intestine. *Proceedings of the National Academy of Sciences* **1997,** *94* (5), 2031-2035.
37. (a) Rowinsky, E.; Eisenhauer, E.; Chaudhry, V.; Arbuck, S.; Donehower, R. In *Clinical toxicities encountered with paclitaxel (Taxol)*, Seminars in oncology, 1993; pp 1-15; (b) Esmaeli, B.; Valero, V.; Ahmadi, M. A.; Booser, D., Canalicular stenosis secondary to docetaxel (taxotere): a newly recognized side effect. *Ophthalmology* **2001,** *108* (5), 994-995.
38. Cai, C., Biodegradable polymer micro-and nanoparticles as protein delivery systems: influence of microparticle morphology and improvement of protein loading capacity of nanoparticles. **2007**.
39. Berlin Jr, C.; May-McCarver, D.; Notterman, D.; Ward, R. M.; Weismann, D.; Wilson, G.; Wilson, J.; Bennett, D.; Hoskins, I.; Kaufman, P., Alternative routes of drug administration-Advantages and disadvantages (subject review). *Pediatrics* **1997,** *100* (1), 143-152.
40. Ting, Y.; Jiang, Y.; Ho, C.-T.; Huang, Q., Common delivery systems for enhancing in vivo bioavailability and biological efficacy of nutraceuticals. *Journal of Functional Foods* **2014,** *7*, 112-128.
41. Xiao, W.; Luo, J.; Jain, T.; Riggs, J. W.; Tseng, H. P.; Henderson, P. T.; Cherry, S. R.; Rowland, D.; Lam, K. S., Biodistribution and pharmacokinetics of a telodendrimer micellar paclitaxel nanoformulation in a mouse xenograft model of ovarian cancer. *Int. J. Nanomedicine* **2012,** *7*, 1587-1597.
42. Shi, C.; Guo, D.; Xiao, K.; Wang, X.; Wang, L.; Luo, J., A drug-specific nanocarrier design for efficient anticancer therapy. *Nature communications* **2015,** *6*.
43. Hunter, C. A.; Sanders, J. K., The nature of. pi.-. pi. interactions. *Journal of the American Chemical Society* **1990,** *112* (14), 5525-5534.
44. (a) Anderson, J. M.; Van Itallie, C. M., Physiology and function of the tight junction. *Cold Spring Harbor Perspectives in Biology* **2009,** *1* (2), a002584; (b) Anderson, J. M., Molecular Structure of Tight Junctions and Their Role in Epithelial Transport. *Physiology* **2001,** *16* (3), 126-130.
45. Frömter, E., The route of passive ion movement through the epithelium ofNecturus gallbladder. *The Journal of membrane biology* **1972,** *8* (1), 259-301.
46. Günzel, D.; Yu, A. S. L., Claudins and the modulation of tight junction permeability. *Physiological Reviews* **2013,** *93* (2), 525-569.

47. Furuse, M.; Fujita, K.; Hiiragi, T.; Fujimoto, K.; Tsukita, S., Claudin-1 and-2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. *The Journal of cell biology* **1998,** *141* (7), 1539-1550.
48. Hewitt, K. J.; Agarwal, R.; Morin, P. J., The claudin gene family: expression in normal and neoplastic tissues. *BMC cancer* **2006,** *6* (1), 186.
49. Krause, G.; Winkler, L.; Mueller, S. L.; Haseloff, R. F.; Piontek, J.; Blasig, I. E., Structure and function of claudins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2008,** *1778* (3), 631-645.
50. Rossa, J.; Protze, J.; Kern, C.; Piontek, A.; Günzel, D.; Krause, G.; Piontek, J., Molecular and structural transmembrane determinants critical for embedding claudin-5 into tight junctions reveal a distinct four-helix bundle arrangement. *Biochemical Journal* **2014,** *464* (1), 49-60.
51. Suzuki, H.; Nishizawa, T.; Tani, K.; Yamazaki, Y.; Tamura, A.; Ishitani, R.; Dohmae, N.; Tsukita, S.; Nureki, O.; Fujiyoshi, Y., Crystal structure of a claudin provides insight into the architecture of tight junctions. *Science* **2014,** *344* (6181), 304-307.
52. Li, J.; Zhuo, M.; Pei, L.; Rajagopal, M.; Yu, A. S. L., Comprehensive Cysteine-Scanning Mutagenesis Reveals Claudin-2 Pore-Lining Residues with Different Intrapore Locations. *Journal of Biological Chemistry* **2014**.
53. Piontek, J.; Winkler, L.; Wolburg, H.; Müller, S. L.; Zuleger, N.; Piehl, C.; Wiesner, B.; Krause, G.; Blasig, I. E., Formation of tight junction: determinants of homophilic interaction between classic claudins. *The FASEB Journal* **2008,** *22* (1), 146-158.
54. Saitoh, Y.; Suzuki, H.; Tani, K.; Nishikawa, K.; Irie, K.; Ogura, Y.; Tamura, A.; Tsukita, S.; Fujiyoshi, Y., Structural insight into tight junction disassembly by Clostridium perfringens enterotoxin. *Science* **2015,** *347* (6223), 775-778.
55. (a) Rossa, J.; Ploeger, C.; Vorreiter, F.; Saleh, T.; Protze, J.; Guenzel, D.; Wolburg, H.; Krause, G.; Piontek, J., Claudin-3 and claudin-5 folding and assembly into the tight junction are controlled by non-conserved residues in TM3 and ECL2 segments. *Journal of Biological Chemistry* **2014**; (b) Gong, Y.; Renigunta, V.; Zhou, Y.; Sunq, A.; Wang, J.; Yang, J.; Renigunta, A.; Baker, L. A.; Hou, J., Biochemical and biophysical analyses of tight junction permeability made of claudin-16 and claudin-19 dimerization. *Molecular biology of the cell* **2015,** *26* (24), 4333-4346; (c) Suzuki, H.; Tani, K.; Tamura, A.; Tsukita, S.; Fujiyoshi, Y., Model for the architecture of claudin-based paracellular ion channels through tight junctions. *Journal of Molecular Biology* **2015,** *427* (2), 291-297.
56. Magrane, M.; Consortium, U., UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011,** *2011*, bar009.
57. Irudayanathan, F. J.; Trasatti, J. P.; Karande, P.; Nangia, S., Molecular Architecture of the Blood Brain Barrier Tight Junction Proteins–A Synergistic Computational and In Vitro Approach. *The Journal of Physical Chemistry B* **2015**.
58. Van Itallie, C. M.; Mitic, L. L.; Anderson, J. M., Claudin-2 forms homodimers

and is a component of a high molecular weight protein complex. *Journal of Biological Chemistry* **2011,** *286* (5), 3442-3450.
59. Zhang, Y., I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **2008,** *9* (1), 40.
60. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **2015,** *10* (6), 845-858.
61. Wiederstein, M.; Sippl, M. J., ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* **2007,** *35* (suppl 2), W407-W410.
62. Zhang, Y., Interplay of I‐TASSER and QUARK for template‐based and ab initio protein structure prediction in CASP10. *Proteins: Structure, Function, and Bioinformatics* **2014,** *82* (S2), 175-187.
63. Abraham, M.; van der Spoel, D.; Lindahl, E.; Hess, B., GROMACS development team (2014) GROMACS User manual version 5.0. 4.
64. Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P., Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation* **2014,** *10* (2), 676-690.
65. Rossa, J.; Ploeger, C.; Vorreiter, F.; Saleh, T.; Protze, J.; Günzel, D.; Wolburg, H.; Krause, G.; Piontek, J., Claudin-3 and claudin-5 protein folding and assembly into the tight junction are controlled by non-conserved residues in the transmembrane 3 (TM3) and extracellular loop 2 (ECL2) segments. *Journal of Biological Chemistry* **2014,** *289* (11), 7641-7653.
66. Van Itallie, C. M.; Anderson, J. M., Claudin interactions in and out of the tight junction. *Tissue Barriers* **2013,** *1* (3), e25247.
67. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM‐GUI: a web‐based graphical user interface for CHARMM. *Journal of computational chemistry* **2008,** *29* (11), 1859-1865.
68. Urry, D. W., Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. *The Journal of Physical Chemistry B* **1997,** *101* (51), 11007-11028.
69. Ipsen, J. H.; Mouritsen, O. G.; Bloom, M., Relationships between lipid membrane area, hydrophobic thickness, and acyl-chain orientational order. The effects of cholesterol. *Biophysical journal* **1990,** *57* (3), 405.
70. Jensen, M. Ø.; Mouritsen, O. G., Lipids do influence protein function—the hydrophobic matching hypothesis revisited. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2004,** *1666* (1), 205-226.
71. Leonenko, Z.; Finot, E.; Ma, H.; Dahms, T.; Cramb, D., Investigation of temperature-induced phase transitions in DOPC and DPPC phospholipid bilayers using temperature-controlled scanning force microscopy. *Biophysical journal* **2004,** *86* (6), 3783-3793.
72. Heitmann, P., A model for sulfhydryl groups in proteins. Hydrophobic

interactions of the cysteine side chain in micelles. *European Journal of Biochemistry* **1968,** *3* (3), 346-350.

7. Bibliography

Xiaoyi Wang

Male, born on Dec 28th, 1990
Address: 126-F Remington Avenue, Syracuse, New York, USA
Tel: +1(315)-396-4863
Email: xwang104@syr.edu

EDUCATION

- Syracuse University
  LC. Smith College of Engineering and Computer Science
  Master of Science in Bioengineering and Biomedical Engineering 1/2014-05/2016
  GPA:3.25/4.00
  900 South Crouse Ave, Syracuse, NY, 13244

- Shanghai Ocean University
  College of Food Science and Technology
  B.S in Biotechnology (Marine Biopharmaceutical Specialty) 09/2009-06/2013
  GPA: 3.15/4.00
  No.999 Huchenghuan Rd, Shanghai, China

RESEARCH EXPERIENCE

- 01/2015-03/2016
  Title: Understanding soft molecular interfaces using multiscale molecular dynamics
  Advisor: **Shikha Nangia** (Assistant Professor of department of Biomedical Engineering and Chemical Engineering, Syracuse University)
  Claudin family of transmembrane proteins play a critical role in tight junctions by regulating paracellular permeability throughout the body including skin, liver, kidney, and lungs. Claudin assembly into tight junctions requires a complex interplay of intracellular and intercellular interactions. Understanding the tight junction assembly is critical in solving diseases associated with claudin mutations. In this work, monomeric structures of seven classic claudins were developed using in silico homology modeling. We have identified fundamental interactions that lead to stable claudin assembly. These results are first of their kind that provide much needed insight into structural and functional attributes of tight junctions.

- 06/2014-12/2015
  Title: Synergistic computational and experimental approaches for optimizing anticancer drug delivery nanocarriers
  Advisor: **Shikha Nangia** (Assistant Professor of department of Biomedical Engineering and Chemical Engineering, Syracuse University)
  Engineered nanocarriers are promising drug delivery vehicles in the field of nanomedicine due to their tunable physiochemical properties. However, designing nanocarriers is challenging due to stringent design requirements such as stability, drug-loading capacity, size distribution, target specificity, and most importantly the nature the encapsulated drug. In this work, we have

developed a collaborative and experimental and multiscale computational modeling approach capable of evaluating three generation of polyethylene glycol-b-dendritic oligomer copolymer (PEG-b-oligomer) called telodendrimer and FDA approved cancer drug Doxorubincin(DOX) macromolecular assemblies (micelles). Results In this work highlight the transition of a nanocarrier shape from spherical to a worm-like micelle with a slight isomeric modification of the nanocarrier building block. These results are in agreement with the experimental observations indicating the success of multiscale computational approaches in optimizing drug delivery nanocarriers based on the nature of the encapsulated drug.

- 12/2012-04/2013
  Title: Verification of thrombosis effect of Panax Notoginseng and speculation of thrombosis mechanism
  Advisor: **Jingya Yang** (Associate Professor of marine biopharmaceutics department, Shanghai Ocean University)
  Panax Notoginseng is a kind of famous Chinese traditional medicine which is widely used for hemostatic and cardiovascular effect. However, the Panax Notoginsengs have functions as both hemostasis and activating blood due to the multiple kinds of effective compounds. In this work, dencichine which is the main effective compounds for hemostatic effect was extracted from raw Panax Notoginseng and applied to in vivo rat. Platelet aggregation and plasma coagulation or rats were investigated and quantified by orbital blood coagulation time, prothrombin time, activated partial thromboplastin time and thrombin time. Mechanism of hemostatic effect for dencichine was discussed and speculated relying on those results.

PUBLICATIONS

1. Wenjuan Jiang, Xiaoyi Wang, Juntao Luo, and Shikha Nangia. "Multiscale approach to investigate self-assembly of telodendrimer based nanocarriers for anticancer drug delivery." In preparation
2. Flaviyan Jerome Irudayanathan, Xiaoyi Wang, Nan Wang, Sarah Rose Willsey, Ian Seddon, Shikha Nangia. "Conformational diversity of classic claudin *cis* interactions - in sights form membrane driven self-assembly simulations" In preparation

INTERNSHIP EXPERIENCE

- 1/23/2013-07/15/2013
  Employer: Shanghai Standard Biotech Co.,ltd. (Shanghai, China)
  1. Analyzed and evaluated 38 kinds of drugs and medical compounds in qualitative and quantitative methods
  2. Developed originally HPLC methodology for 2 kinds of traditional medical compounds
  3. Testified anticancer effects of medical compounds by Elisa, SRB and Western blot
  4. Developed Elisa kit for internal use

- 07/2012-09/2012
  Employer: Second Military Medical University (Shanghai, China)
  1. Participated in experimental design for a new kind of azole nitrogen class of anti-fungal compound
  2. Synthesized and separated the target product by silica column chromatography
  3. Testified the structural formula by using 1H NMR, and 13C NMR

- 07/2011-09/2011
  Employer: Shanghai Food and Drug Administration (Shanghai, China)
  1. Investigated products in pharmacy, hospital, and mall, and checked their legality
  2. Rapid inspected cosmetics, health care products, data collection and entry
  3. Collaborated staff to generate legal documents, records, and made backup

AWARDS & HONORS

- 06/2013 Shanghai Ocean University Outstanding Graduates (5%)
- 09/2009 – 01/2013 Shanghai Ocean University First Prize Scholarship 2 times (5%), Second Prize Scholarship 3 times (10%), Third Prize Scholarship 2 times (20%) (7 times/7 Semesters in total
- 09/2012– 06/2013 Outstanding League Cadre (2%)
- 09/2011 – 06/2012 Pacemaker to Merit Student (3%)
- 2012 'Tingya' Individual Private Scholarship, Enterprise Scholarship (1%)